

PH.D. THESIS

Stochastic Models for the Estimation of Airport Arrival Capacity Distributions

by T. Inniss

Advisor: M. Ball, B. Golden, P. Schonfeld, S. Lele, P. Smith

NEXTOR PhD 2001-1
(ISR PhD 2001-3)

NEXTOR

National Center of Excellence in Aviation Operations Research

The National Center of Excellence in Aviation Operations Research (NEXTOR) is a joint university, industry and Federal Aviation Administration research organization. The center is supported by FAA research grant number 96-C-001.

Web site <http://www.isr.umd.edu/NEXTOR/>

STOCHASTIC MODELS FOR THE ESTIMATION
OF AIRPORT ARRIVAL CAPACITY DISTRIBUTIONS

by

Tasha R. Inniss

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland at College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
July 2000

National Center of Excellence for Aviation Operations Research
NEXTOR Report #T-01-1

Advisory Committee:

Professor Michael O. Ball, Chairman/Advisor
Professor Bruce Golden
Professor Paul Schonfeld
Professor Shreevardhan Lele
Professor Paul J. Smith

ABSTRACT

Title of Dissertation: STOCHASTIC MODELS FOR
THE ESTIMATION OF AIRPORT
ARRIVAL CAPACITY DISTRIBUTIONS

Tasha R. Inniss, Doctor of Philosophy, 2000

Dissertation directed by: Professor Michael O. Ball
Decision and Information Technologies,
Robert H. Smith School of Business

In this dissertation, statistical and integer programming methods are used to calibrate models to estimate airport arrival capacity distributions. These distributions are an essential input to decision models used to regulate flow into congested airports when demand for arrival resources exceeds the available capacity. The techniques developed make contributions to the body of knowledge on air traffic flow management. On a more general level, the approach developed can be viewed as a clustering technique that maintains the time order of imbedded time series data.

During instances of capacity-demand imbalances, efficient planning and decision-making is contingent upon the “goodness” of the models that estimate airport capacity over time. Airport capacities are subject to substantial uncertainty as

they depend on stochastic weather conditions. The models developed in this thesis are required inputs into a class of stochastic ground holding models that determine the amount of ground delay to assign to incoming flights to balance assigned ground delay and expected airborne delay optimally while minimizing total delay. The models are judged by the amount of total weighted delay incurred in comparison to the amount of total weighted delay that would have been realized under current operational procedures. (Airborne delay is weighted more heavily than ground delay). Based on comparisons between the decision models that employ the estimates developed in this thesis and current models, the results of this thesis reduce the total amount of weighted delay. Another contribution is the development of a new, simple decision model that more accurately estimates the amount of delay incurred in a ground delay program as it dynamically changes.

The statistical models calibrated in this thesis use empirical historical data to generate (arrival capacity) probability distribution functions (vectors of capacity scenarios). In the case considered in this thesis, the capacity scenarios (one-parameter arrival capacity distributions) are used to model morning fog at San Francisco's International Airport. To determine seasonal capacity probabilistic distribution functions (distributions that vary in time), a set partitioning integer programming model is utilized.

© Copyright by
Tasha R. Inniss
2000

PREFACE

This report documents research supported by the National Center of Excellence for Aviation Operations Research, under Federal Aviation Administration Research Grant Number 96-C-001. This document has not been reviewed by the Federal Aviation Administration (FAA). Any opinions expressed herein do not necessarily reflect those of the FAA or the U.S. Department of Transportation.

DEDICATION

I dedicate this entire endeavor and achievement to my very loving and supportive family, especially my grandfather, Clarence J. Inniss, who inspired me to be a mathematician, and my mother, Dr. Leslie B. Inniss, who is the embodiment of strength and goodness.

ACKNOWLEDGEMENTS

I would like to begin by thanking my dissertation advisor, Dr. Michael O. Ball, whose eternal patience, tireless guidance and infinite insight made this all a reality. A very special thanks to my unofficial advisor and mentor, Dr. Paul J. Smith, for all the long hours he spent sharing his statistical genius with me. To my dissertation advisory committee, specifically Dr. Shreevardhan Lele, Dr. Paul Schonfeld and Dr. Bruce Golden, thank you for taking time to read my dissertation, for offering valuable suggestions and for serving on my committee.

To the CDM community, especially Forrest Terral, Jim Wetherly, Jill Charlton, Dan Citrenbaum and Joe Rather at the Federal Aviation Administration; and Midori Tanino, Lara Shisler, and Robert Hoffman at Metron, Inc., I am grateful for your invaluable contributions and insight into my research. I appreciate your giving generously of your time and expertise to help me obtain data necessary for my thesis and to gain a complete understanding of air traffic flow management and procedures at the ATCSCC. A special thanks to Dr. F. Wesley Wilson and David A. Clark at MIT's Lincoln Laboratory for providing weather data and sharing knowledge about marine stratus conditions

at SFO. I appreciate your giving me permission to reproduce graphics from your articles in my thesis.

The support and understanding of family, friends and my colleagues in the mathematical community gave me a tremendous spiritual boost that helped me achieve this goal. I received a solid foundation of love and constant encouragement from my birth family and a strong academic foundation from my family at Xavier University of Louisiana. To my brother Enos, the most important person to me: Thank you for being a strong, dedicated, encouraging and loving brother. You have always been my number one supporter and always know how to make me laugh even in my darkest hours. I would like to give special recognition to my very best friends, Brian A. Williams, Atiya N. Hoyer, and Dionne L. Price; To Bri, my soul's inspiration: From you I derive strength because you always know my unspoken needs and allay my unspoken fears. To Atiya, my voice of reason: Thank you for your refreshing honesty and unconditional friendship. You know best how to keep me level-headed. To Dionne, my dissertation buddy: Thank you for keeping me on track, focused and motivated. You could empathize most and I am grateful for your daily words of encouragement. To my wonderful mentors, Dr. Raymond Johnson and Dr. James Turner, who are guiding and preparing me to be a professional mathematician: Thank you for your unending advice and mentorship, your unwavering belief in me, and your unceasing encouragement of my endeavors.

My entire graduate school tenure was generously funded by the David

and Lucile Packard Foundation and the Southern Regional Education Board (dissertation fellowship). I not only appreciate the financial support, but also all the encouragement and inspiration you provided. This research was supported in part by the Federal Aviation Administration under the National Center of Excellence for Aviation Operations Research (NEXTOR).

TABLE OF CONTENTS

List of Tables	xiii
List of Figures	xv
1 Introduction	1
1.1 Ground Holding Procedures	2
1.2 Collaborative Decision Making (CDM) and GDP Enhancements .	4
1.3 Motivation for Problem Studied	7
1.4 Contents and Contributions of the Thesis	10
2 Background and Literature Review	13
2.1 Discussion of Airport Capacity	13
2.2 Ground Holding Models	16
2.3 Integer Programming	21
2.3.1 Totally Unimodular and Interval Matrices	22
2.3.2 Set Covering and Set Partitioning	23
2.4 Statistical Background	24
2.4.1 Probability and Empirical Distribution Functions	24
2.4.2 Time Series	28
2.4.3 Analysis of Variance (ANOVA)	30

3	Description of Data: Sources and Preparation	35
3.1	Ground Delay Program (GDP) Logs	36
3.2	Surface Airways Hourly Weather Data	39
3.3	Merging GDP Logs and Weather Data	41
3.4	Aggregate Demand Lists	41
4	Airport Arrival Capacity Distributions (Capacity Scenarios)	43
4.1	Types of Arrival Capacity Distributions	43
4.2	Estimating 1-Parameter ACDs at SFO	48
4.3	Determining Distribution of 1-Parameter ACDs	49
5	Determining Capacity Probabilistic Distribution Functions (CPDFs) that Vary in Time	56
5.1	Daily CPDFs	57
5.2	Monthly CPDFs	61
5.3	Seasonal CPDFs	62
5.3.1	Detecting Seasonal Trends	62
5.3.2	Model Formulation	65
5.3.3	Determining Seasonal Clusters using Average GDP Durations	71
5.3.4	Determining Seasonal Clusters using Empirical Distribu- tion Functions	77
5.3.5	Post Analysis for Evaluating Sets of Seasons	79
5.3.6	Seasonal “Clustering” Technique	89
5.4	Extension to CPDFs of 2-Parameter ACDs	92
6	Decision Models and Evaluation of ACD Models	94
6.1	The Hoffman-Rifkin Static Stochastic Ground Holding Model . . .	94

6.2	Adjusting Assigned Ground Delay for Dynamically Changing GDPs	96
6.2.1	Shortened Reduced Capacity Due to Canceled GDP	97
6.2.2	Lengthened Reduced Capacity Due to Revised GDP	103
6.3	Analysis and Computational Results	105
6.3.1	H-R Results (M_PAAR) for Seasonal CPDFs	105
6.3.2	Procedure for Comparing Planned and Actual Capacity Scenarios	107
6.3.3	Comparisons of H-R Results with Command Center Plans	115
6.4	Proposed General Decision Model	117
6.5	Benefits Analysis and Future Vision of CDM	118
6.5.1	Increase in GDP Effectiveness under CDM	118
6.5.2	Incremental GDPs	122
7	Conclusions	124
7.1	Contributions of Thesis	124
7.2	Directions for Further Research	125
A	Glossary	127
B	MIT's Lincoln Laboratory Graphs and Chapter 5 Tables	131
C	Chapter 5 Graphs	136
D	Comparison of Results of H-R Model (M_PAAR) and Command Center Plan (CC_PAAR)	145
	References	145

LIST OF TABLES

2.1	AAR Chart for SFO (Courtesy of ATCSCC)	14
2.2	ANOVA Table for One-Way Layout	32
3.1	Data from GDP Logs	36
3.2	Number of GDPs at SFO	37
3.3	Data from Surface Airways Hourly	40
3.4	Codes for Weather Data in Surface Airways Hourly	40
3.5	Data fields in the Aggregate Demand Lists	42
5.1	Seasonal Clustering Criteria (Cost Functions)	73
5.2	Set Covering Solutions of GDP Seasons (n=60)	75
5.3	ACD Probabilities for Frequency and Smoothed Histograms	77
5.4	Set Covering/Partitioning Solutions using KS statistics (n=60) . .	80
5.5	Results of ANOVA Multiple Comparisons Test	85
5.6	F-Test Results for OLS and WLS	86
5.7	Mean Square Ratios of Weather Seasons	88
5.8	Mean Square Ratios of “Standard” Seasons	89
6.1	Filtering Criteria	102
6.2	Actual Percentages of GD Recovered in a Canceled GDP	103
6.3	Average Percentages of GD Recovered in a Canceled GDP	104

6.4	Capacity Scenarios (Inputs into the H-R Model)	107
6.5	Results of the H-R Model (Number of Hours of Reduced Capacity)	107
6.6	Total Weighted Delay of H-R Plans Vs Command Center Plans .	115
B.1	Set Covering and Partitioning Solutions of GDP Seasons(n=133) .	134
B.2	Set Covering/Partitioning Solutions using KS statistics (n=133) .	135

LIST OF FIGURES

1.1	Proposed GDP-E Concept of Operations	8
2.1	Runway Layout at SFO (Courtesy of the ATCSCC)	15
3.1	Percentage of GDPs Canceled at SFO	38
3.2	Percentage of GDPs Implemented During Morning Hours at SFO	38
4.1	Form of the General Arrival Capacity Distribution	45
4.2	Form of the 2-Level Arrival Capacity Distribution	45
4.3	Form of 2-Parameter Arrival Capacity Distribution	47
4.4	Form of 1-Parameter Arrival Capacity Distribution	47
4.5	Overall Frequency Histogram for SFO Morning GDP Data	50
4.6	Smoothed Histogram for SFO Morning GDP Data	51
4.7	Relative Frequency Histogram for Duration of IFR Conditions . .	52
4.8	“Conditional” Distribution of Durations of IFR Conditions	53
5.1	Relative Frequency Histogram (Uniform Weighting) for January 1st	58
5.2	Relative Frequency Histogram (Uniform Weighting) for January 2nd	59
5.3	Relative Frequency Histogram (Uniform Weighting) for January 3rd	60
5.4	Time Series Plot of Average GDP Durations	63
5.5	Average Centered Moving Averages of GDP Durations	64

5.6	Average Centered Moving Averages of Durations of IFR Conditions	64
5.7	Network Representation of Seasonal Clustering Problem	67
5.8	Relationship Between Mean and Variance of GDP Duration	72
5.9	Average of Monthly Average GDP Durations over 1995, 1996, 1997	76
5.10	Time Plot of Residuals of Daily GDP Durations	81
5.11	QQ Plot of Studentized Residuals of Daily GDP Durations	82
5.12	Time Plot of Residuals of Monthly GDP Average Durations	83
5.13	QQ Plot of Studentized Residuals of Monthly GDP Average Du- rations	84
5.14	Perspective on Seasonal “Clustering”seasonalclustering	91
6.1	Ground Delay Available for Recovery (Recoverable GD)	98
6.2	Plot of Percentage of Ground Delay Recovered as a Function of Dif- ference of (Controlled) Departure Times from Cancellation Times of GDPs	100
6.3	Filtered Percentage of GD Recovered for Percentages [-10,110] . .	100
6.4	Percentage of GD Recovered for Assigned GD < 2 Hours	101
6.5	Percentage of GD Recovered for Flights Whose (Controlled) De- parture Times Are Less Than 3 Hours After Cancellation Times of GDPs	101
6.6	Plot of Percentage GD Recovered After All Filtering	102
6.7	Additional GD under Revised GDP	104
6.8	Conditional Frequency Histogram for “Heavy Fog” Season	108
6.9	Conditional Frequency Histogram for “Rainy” Season	109
6.10	Conditional Frequency Histogram for “Summer Weather” Season .	110
6.11	Percentage of Outcomes of 1998 Morning GDPs	116

6.12	Percentage of GDPs Canceled	119
6.13	Percentage of GDP Cancellations Near Start	120
B.1	IFR Capacity at SFO	132
B.2	Frequency of IFR events at Major US Airports	133
C.1	Relative Frequency Histogram for Winter GDP Season	137
C.2	Relative Frequency Histogram for Spring GDP Season	138
C.3	Relative Frequency Histogram for Summer GDP Season	139
C.4	Relative Frequency Histogram for Fall GDP Season	140
C.5	Smoothed Histogram for Winter GDP Season	141
C.6	Smoothed Histogram for Spring GDP Season	142
C.7	Smoothed Histogram for Summer GDP Season	143
C.8	Smoothed Histogram for Fall GDP Season	144

Chapter 1

Introduction

Domestic air traffic has greatly increased over the last 10 years and is predicted to continue to increase at a rate of 3 to 5% over the next 15 years.¹. With this great increase in air traffic comes a large increase in the demand for airspace and airport resources. Unfortunately, airspace and airport capacities are not increasing at a rate necessary to meet this rising demand. It is vital that new methodologies and tools be developed to address the inevitable rise in congestion. Because of this surge in air traffic and the limited capacity of airports, air traffic flow management is becoming an increasingly difficult task.

When an airport's capacity is reduced during "peak demand hours", demand for an airport's resources exceeds the capacity at which the airport can meet this demand. This is known as a capacity-demand imbalance. Demand refers to the number of flights scheduled to arrive or depart in a given time period (rate of flight arrivals or departures). Capacity is the maximum number of flight arrivals or departures in a given time period. To address an arrival capacity-demand imbalance, air traffic flow managers in the United States institute ground holding

¹From the "CDM GDP Training Website" on Metron's Homepage:
<http://www.metsci.com/cdm/newmember.html#>

procedures to align the demand with the available capacity.

1.1 Ground Holding Procedures

During instances of capacity-demand imbalances, air traffic flow management (TFM) in an efficient and safe manner is of premier importance. Any given airspace is composed of flight paths, control facilities, sectors and airports. The overall goal of TFM is to strategically plan and manage entire flows of air traffic, provide the greatest and most equitable access to airspace resources, mitigate congestion effects from severe weather and ensure the overall efficiency of the system without compromising safety. In the United States' National Airspace System (NAS), there are 21,000 daily commercial flights that are monitored and controlled by 21 Air Route Traffic Control Centers (ARTCCs), 462 airport towers and 197 Terminal Radar Approach Control Facilities (TRACONs). footnote- See Metron's homepage [23], www.metsci.com/cdm. The entire United States airspace is monitored by a central Federal Administration Agency (FAA) facility known as the Air Traffic Control System Command Center (ATCSCC) located in Herndon, Virginia. A fundamental capability of all TFM centers is the ability to monitor airspace for potential capacity-demand imbalances.

Managing traffic flow during capacity-demand imbalances is known as the traffic flow management problem (TFMP). The TFMP has become increasingly more important and difficult as the amount of air traffic has increased over the years. The seriousness of this problem is evidenced by a steady increase in delays. Ground holding procedures are a principal tool used to address the TFMP.

The two main ground holding procedures employed are ground stops and ground delay programs (GDPs). A ground stop is an extreme FAA initiative

taken when arrival capacity drastically drops suddenly or when it is greatly underestimated. In a ground stop, flights are held on the ground at their departure airports until it is determined that the capacity-demand imbalance has abated. The most widely used ground holding procedure is a GDP. In a GDP, flights in a certain time window or at a certain distance away from a congested airport are assigned delay to be taken on the ground until a time that they can safely land at their terminal airports with little to no airborne delay. Currently during a GDP, major airports submit to the ATCSCC their forecasted capacity on an hour-by-hour basis. The specialists view projected demand for each of these airports to determine if there will exist considerable capacity-demand imbalances. If there will be a severe capacity-demand imbalance, specialists institute a national GDP for the airports impacted. An advisory is sent out to airlines notifying them of the plan to institute a GDP. The advisory contains the affected airports, the planned start time of the GDP and the capacity that will be used to generate the GDP. Airlines then have the option of responding to the advisory with a submission of schedule changes in the form of delays or cancellations of flights. The specialists at the ATCSCC then reevaluates the demand at the affected airports to determine if a GDP is still needed. If one is still needed, flights originally scheduled to land at a congested airport during the time of the GDP are given delay to be taken on the ground at their departure airports.

It should be noted that ground stops and GDPs are the domain of traffic flow managers. In contrast, air traffic controllers (ATC) are concerned with the control of a single flight throughout its whole course to ensure safety. ATC is centered around shorter-term individual flight management, whereas TFM is concerned with aggregates of flights throughout the whole airspace. The increasing severity

of the TFMP in the United States has led to the development of a new TFM paradigm, Collaborative Decision Making (CDM).

1.2 Collaborative Decision Making (CDM) and GDP Enhancements

Collaborative Decision Making (CDM), now known as Collaborative Traffic Flow Management (CTFM), was motivated by a need for increased information sharing and distributed decision-making. There was a desire to shift from a central planning paradigm to a collaborative TFM paradigm in which airlines, through their airline operational control centers (AOCs), would have more control, more flexibility and more input into the air traffic flow management decision-making processes. The philosophy of CDM is that with increased data exchange and collaboration comes better and more effective decisions on the part of the traffic flow managers. CDM goes hand-in-hand with the ATC concept of Free Flight, an architecture in which more responsibility for flight maneuvering and aircraft separation is given to the aircraft and pilot.

CDM was initially conceived in the fall of 1993 in an experiment known as FADE (FAA/Airline Data Exchange). FADE was conducted to ascertain the impact of airlines' sending in updated scheduling information. In the summer of 1994, a human-in-the-loop exercise was conducted at the ATCSCC to ascertain the impact of dynamic schedule updates on the specialists' decision-making processes. It was determined from both experiments that dynamic schedule updates decreased airline delay and had a positive impact on the decision-making processes of the specialists at the ATCSCC. CDM evolved out of FADE and was

officially formed in 1995.²

In the spring of 1995, a “Role and Responsibilities” guideline for CDM was drafted by the airlines and agreed to by the ATCSCC. In the guideline, it is stated that the role of the ATCSCC is to monitor the NAS for constraints, make the constraints known to the NAS users and work in cooperation with the NAS users to develop solutions. The role of the NAS users is to notify the ATCSCC of all schedule updates and changes based on the agreed upon solutions. Basically, the ATCSCC would look for constraints, ration limited arrival resources and allow the NAS users to plan and act according to their own economic objectives and user-defined goals. In order to facilitate the above roles, Metron, Inc., a scientific consulting firm, developed a decision-support tool known as the Flight Schedule Monitor (FSM). Through FSM, both the ATCSCC and the NAS users can view the predicted demand, capacity and all arrival information at a given airport and can run “what if” scenarios to determine the impact of certain decisions. In “what if” scenarios, different GDP parameters such as duration of a GDP and scope of a GDP are input to ascertain the resulting amount of delay and number of flights affected by the GDP. One of the most important aspects of CDM is common situational awareness of all constraints in the NAS. This is achieved through FSM.

In the summer of 1996, the FAA commissioned the National Center of Excellence for Aviation Operations Research (NEXTOR) to aid in addressing and developing solutions to air traffic flow management problems. NEXTOR is a consortium of four universities, the University of Maryland, the Massachusetts

²Most of the information in this section can be found at Metron’s website: www.metsci.com/cdm

Institute of Technology, the University of California at Berkeley, and the Virginia Polytechnic University, certain affiliate universities and 20 industry partners including Metron, Inc., Honeywell and Boeing.³ The CDM working group is comprised of representatives from the FAA, the ATCSCC, the 33 participating airlines and subcarriers, industry and the research and development community including NEXTOR. The CDM working group meets monthly to discuss the progress of the CDM paradigm and is currently expanding their efforts from GDP planning to tackling problems within the enroute airspace. The new challenge is to apply and implement principles and processes of CDM for route planning. This process is known as Collaborative Routing (CR).

In this dissertation, the focus is on CDM efforts during the planning and implementing of GDPs. The CDM working group focused initially on improving GDP procedures, which are now known as GDP Enhancements (GDP-E). GDP-E went into initial prototype operations in January 1998 at San Francisco International Airport (SFO) and Newark Airport (EWR) and full prototype operations in September 1998 at all airports. Under new CDM GDP-E procedures, demand is more accurately projected based on dynamic schedule updates from the airlines. Currently, airlines have an incentive to send in schedule updates since delays and control times are now given based on original scheduled times of arrival versus estimated times of arrival. This procedure of assigning limited arrival resources based on original scheduled arrival times is known as Ration by Schedule (RBS). RBS is based on fair allocation principles, which have eliminated implicit penalties to the airlines for providing up-to-date information. Other benefits to GDPs

³For more information about NEXTOR, visit www.nextor.umd.edu or www.its.berkeley.edu/nextor

under CDM have been derived from the development of the Compression Algorithm that is used for inter-airline swapping of arrival slots that have been vacated due to delays and cancellations and the ability of the specialists at the ATCSCC to make dynamic revisions to GDP parameters. For additional benefits, see Ball et al [4] and [24].

1.3 Motivation for Problem Studied

In this dissertation, models are developed to estimate arrival capacity distribution functions (vectors of capacity scenarios) that are used as inputs into a class of stochastic ground holding models, which describe airport capacity with probability distribution functions. Prior to now, there were no methods for determining capacity probability distribution functions. The use of random variables to describe capacity allow for consideration of stochastic weather conditions that is a reflection of real, every day conditions. Thus, arrival capacity distributions are better representations of the true nature of arrival capacity than deterministic values set at the beginning of the day.

The estimation of arrival capacity distributions is most crucial during those times when the demand for an airport's arrival resources exceed the available arrival capacity due to inclement weather conditions. CDM has been shown to have made significant positive impacts on GDPs during instances of capacity-demand imbalances. The stochastic models developed in this thesis will be an integral part of the GDP process under CDM. After careful examination of the procedures taken by the specialists during a GDP, the flow chart in Figure 1.1 was developed and depicts a proposed method of operations that is completely in line with current CDM procedures.

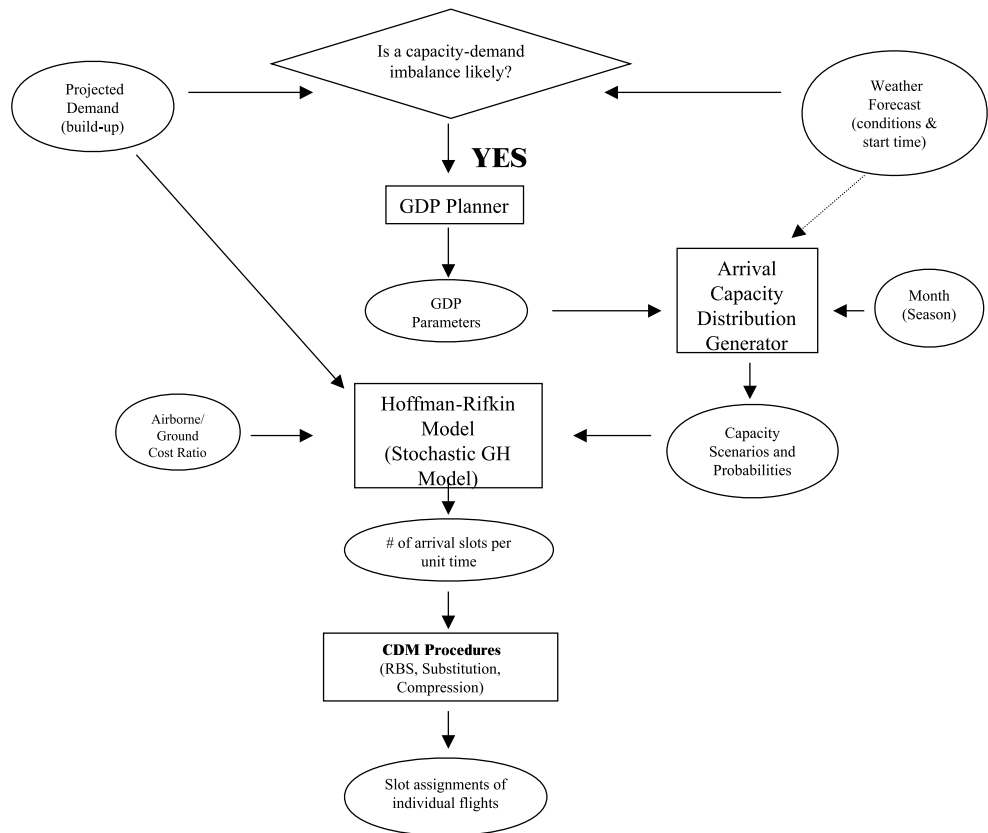


Figure 1.1: Proposed GDP-E Concept of Operations

The flow chart, which will be referred to as the (Proposed) GDP-E Concept of Operations, starts with the assessing of whether a capacity-demand imbalance exists or is likely. This assessment is based on weather conditions such as ceiling and visibility and the predicted arrivals or projected demand at an impacted airport. If a capacity-demand imbalance is likely, then the specialists plan a GDP. The GDP parameters such as the planned airport acceptance rate (PAAR) and the duration of the GDP are used as inputs into the “arrival capacity distribution generator”, which is based on the models in this dissertation. The other required input is the time of year (day, month or season), which determines which capacity probabilistic distribution will be used. The arrival capacity distribution generator results in capacity scenarios and associated probabilities that will be used as inputs into a stochastic ground holding model along with the aggregate projected demand and the airborne to ground delay cost ratio. The output from the ground holding model are the number of arrival slots to make available in each time period or the number of flights that should be assigned ground delay in each time period. The CDM procedures are then used to determine the individual flights that should be assigned ground delay.

Capacity scenarios and associated probabilities are crucial inputs and necessary for any stochastic ground holding model. Due to the facts that no methods have been previously developed to determine capacity scenarios and the models developed in this thesis can be used in conjunction with current ATCSCC CDM procedures, this thesis makes valuable contributions to air traffic flow management.

1.4 Contents and Contributions of the Thesis

The focus of this dissertation is the calibration of models to estimate arrival capacity distributions and the development of cost models that comparatively evaluate a proposed planned airport acceptance rate (PAAR) relative with an actual AAR. The contents and contributions of each subsequent chapter are listed below.

Chapter 2: This chapter gives a background and description of airport capacity along with a literature review of ground holding models, integer programming and the statistical techniques used in the thesis.

Chapter 3: This chapter provides descriptions of the data sources and definitions of the data fields used to calibrate the statistical models of the thesis. There are 3 main data sources: Ground Delay Programs' Logs (GDP data), Surface Airways Hourly Weather Data and the Aggregate Demand Lists. The data from the latter source is used in Chapter 6 to evaluate the statistical models and develop a general decision model to determine the appropriate PAAR. The former two sources are the main sources used to calibrate the statistical models. The models based on GDP data could be used as input into the Hoffman-Rifkin model when weather data is not available. The GDP data is used to perform analysis on the impacts of CDM on GDPs (Section 6.5.1). The data most appropriate to develop the statistical models is a combination of the GDP data and the weather data. The type of model needed is a conditional distribution, a distribution of the length of inclement weather conditions given that a GDP is planned. Since either inclement weather conditions or present or they are not, the weather data does not provide a probability of having 0 hours of inclement weather conditions. The GDP data is used to provide this probability. The resulting distribution

gives capacity scenarios and associated probabilities that are entered into the Hoffman-Rifkin model when both types of data are available.

Chapter 4: A capacity scenario is a distribution of arrival capacity over a given period of time. Depending on the airport, the runway configuration, the predicted demand and the severity of the weather conditions, a capacity scenario can have a very complicated structure. This chapter presents four types of representative capacity scenarios that are referred to as arrival capacity distributions (ACDs). It should be noted that the types of ACDs presented are not exhaustive and do not represent ALL possible types of capacity scenarios, but do capture the majority of scenarios at most major airports. It was discovered that the scenarios with relatively simple structures are representative of capacity scenarios at a broad range of airports. Since a capacity scenario is one realization of capacity, a method is presented for determining a vector or probabilistic distribution of capacity scenarios. This distribution is called a capacity probabilistic distribution function (CPDF). The chapter ends with the derivation of the (overall) conditional CPDF using GDP data and weather data.

Chapter 5: This chapter begins with a discussion of the types of CPDFs that can vary over time depending on the underlying changing weather conditions. These types include daily CPDFs, monthly CPDFs and seasonal CPDFs. The primary focus of the chapter is on the seasonal CPDFs. The determination of the “best” set of seasons is thoroughly explored. Seasonal clusters are determined by a set partitioning algorithm whose objective is to minimize the sum of the costs of each season. The cost function for the algorithm is based on the “difference” between a season’s distribution and the distributions of the months contained in the season. Since there are many ways to calculate this difference, the various

cost functions are presented and resulting seasons are evaluated based on the amount of homogeneity within a season and the variability between seasons. The techniques presented gave rise to a type of clustering technique that determined clusters while maintaining the time order of the data.

Chapter 6: The Hoffman-Rifkin static stochastic ground holding model is described in greater detail than in Chapter 2. It is shown that the Hoffman-Rifkin model does not accurately model GDP dynamics. Therefore, cost models are derived to estimate the amount of total weighted delay for a given planned capacity scenario. The results of the Hoffman-Rifkin model that use the CPDFs as input with the added modifications are evaluated against the plans of the ATCSCC on actual GDPs in 1998. The modified decision model results in a reduction in total weighted delay during a GDP. An algorithm is given to determine the best PAAR given any resulting actual scenario. This decision model incorporates the dynamic changes in the amount of assigned ground delay due to a canceled GDP or revised GDP.

Chapter 7: Conclusions and suggestions for future research are provided.

Chapter 2

Background and Literature Review

2.1 Discussion of Airport Capacity

The effectiveness of a GDP is contingent upon an accurate demand profile and a true representation of an airport's available capacity during inclement weather conditions. CDM procedures have contributed greatly to an increase in the accuracy of aggregate demand at airports, but have done little to determine the actual available capacity at congested airports. Thus, the focus of the dissertation is the development of models to capture the true nature and behavior of airport (arrival) capacity.

An airport's capacity or airport acceptance rate (AAR) is directly related to weather conditions through an airport's runway configuration and its landing procedures. Inclement weather conditions at an airport are used to determine which runway configurations to institute and which landing procedures to implement. There are 2 major types of landing procedures: Instrument Flight Rules (IFR) and Visual Flight Rules (VFR). IFR is required when a (cloud) ceiling of less than 1000 feet or a visibility of less than 3 miles exists. VFR refers to weather conditions that have a ceiling that exceeds 1000 feet and a visibility that exceeds

Land	Depart	IFR	VFR	VAPS
28L 28R	1L 1R	30	45	60 (daylight); 50 (non-daylight)
28L 28R	28L 28R	30	45	45
28L or 28R	1L 1R	30	N/A	30
28L 28R	1L or 1R	30	45	45
1L 1R	1L 1R	30	N/A	30
19L 19R	10L 19R	27-30	N/A	45
19L 19R	19L 19R	25-30	N/A	42
19L or 19R	10L 10R	27-30	N/A	30
19L 19R	10L or 10R	27-30	N/A	45
10L 10R	10L 10R	27-30	N/A	37
Any Single	Runway	27	N/A	27

Table 2.1: AAR Chart for SFO (Courtesy of ATCSCC)

3 miles.

Under VFR conditions at SFO, aircraft normally arrive from the northwest in dual side-by-side approaches on runways 28L and 28R. See Figure 2.1 for runway layout at SFO. When IFR conditions exist, the AAR is reduced because the landing of aircraft in pairs on the two closely spaced parallel runways is considered unsafe. The combination of the runway configurations and the landing procedures determines an airport's operational capacity or AAR. Table 2.1 lists the capacities or AARs for the various combinations of runway configurations and landing procedures at SFO. VAPS is an acronym for visual approaches and has the same conditions as VFR with the addition of a ceiling that exceeds 3500 feet and a visibility that exceeds 7 miles at the San Mateo Bridge for SFO.

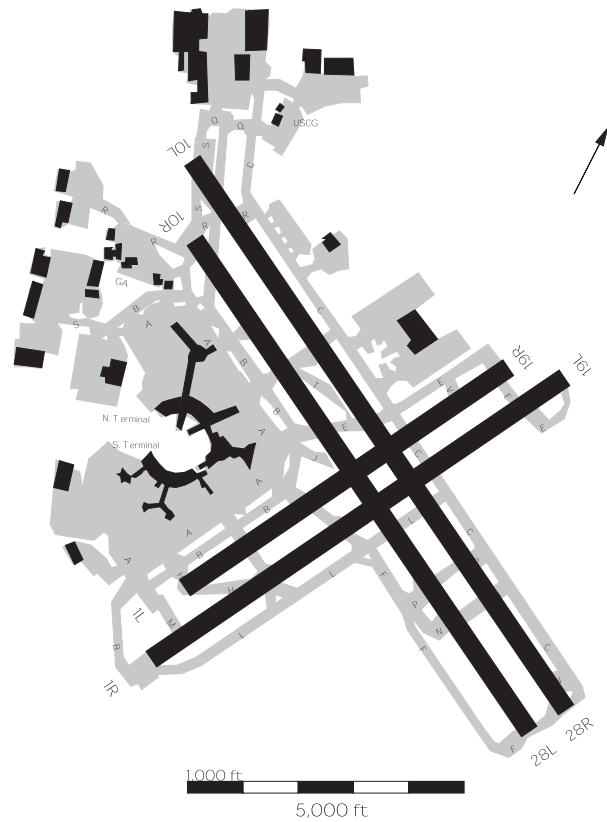


Figure 2.1: Runway Layout at SFO (Courtesy of the ATCSCC)

Airport capacity is comprised of two interdependent capacities, the arrival capacity and the departure capacity. The determination of an airport's capacity is a difficult task because weather conditions, runway configurations, arrival/departure ratios and the fleet (aircraft type) mix must all be considered. Eugene Gilbo, from the Volpe Transportation Systems Center, proposes methods in [16] and [17] for optimizing airport capacity by considering the complex relationship between arrival and departure capacities through an arrival/departure capacity curve. Hall [19] expands Gilbo's work by developing collaborative methods for allocating arrival and departure capacities. Hall's work extends the CDM GDP-E procedure and concepts to this more complex setting. Though it is recognized that additional efficiencies can potentially be gained by considering both arrival and departure capacities simultaneously, arrival capacities and estimating arrival capacity distributions for a GDP will be the focus of this thesis.

2.2 Ground Holding Models

The effective assignment of delay to flights during a GDP is a crucial element to the effectiveness and fairness of a GDP. Fairness of a GDP refers to equitable allocation of delay to each airline. There is a constant hedging between conservative policies of assigning more ground delay that could lead to the underutilization of arrival resources and the liberal policies of assigning less ground delay that could lead to more costly airborne holding delays. Thus, the ground holding problem (GHP) seeks to determine an optimal balance between these policies for assigning delay in a GDP.

The GHP was first discussed and described by Odoni in 1987 [31]. Odoni referred to the deterministic GHP as the flow management problem in which

travel times and capacities are deterministic, there exists a discrete time horizon and the only capacitated element is the arrival airport.

There are three main assumptions required by the ground holding model, which are based on the assumptions of the flow management problem. The assumptions are (1) a discrete time horizon, (2) deterministic demand and (3) deterministic capacity. The following is an integer programming formulation of the GHP:

$$\begin{aligned}
& \text{Minimize} && \sum_{f \in F} \sum_{t=1}^T X_{ft} C_{ft} (t - a_f)^\sigma \\
& \text{subject to} && \sum_{t=a_f}^T X_{ft} = 1 \\
& && \sum_f X_{ft} \leq b_t \\
& && 0 \leq X_{ft} \leq 1
\end{aligned}$$

where X_{ft} is a binary variable which takes value 1 when flight f is assigned to time period t and 0 otherwise, C_{ft} is a cost factor associated with assigning flight f to time period t , T is the number of time periods t , a_f is the scheduled arrival time period of flight f , σ is a super-linear growth parameter used to penalize increasing tardiness from the scheduled arrival time period and b_t is the arrival capacity for time period t .

There have been many mathematical programming models developed to solve different versions of the GHP. There are different terms in the literature used to define and distinguish the different versions of the GHP. “Static” means that ground and airborne holds are decided once at the beginning of the day. “Dynamic” refers to models that allow for periodic updates as better weather and

capacity information become available. “Deterministic” means that airport capacities can be forecasted exactly where as “stochastic or probabilistic” means that the capacities are random variables described through probability distribution functions. This thesis is motivated by the requirement of stochastic versions of the GHP for capacity probability distribution functions.

In 1987, Andreatta and Romanin-Jacur [2] developed a dynamic programming algorithm for the single-airport static stochastic GHP for at most one time period. This was the first paper written that developed an algorithmic approach to determining the amount of ground delay to assign to flights bound for a congested airport. The authors considered a single destination airport and n flights bound for this airport. The dynamic program resulted in an optimal delay strategy that minimized total expected delay for the n flights. The model in this paper is a static, stochastic version of the GHP because it is assumed that airport capacity information is known at the beginning of the day and is summarized using a random variable:

“Airport capacity can be summarized by a random variable K_t that takes on $0, 1, \dots, n$ with probability $p(0), p(1), \dots, p(n)$.”

In his thesis, Terrab [40] developed an efficient algorithm to solve the single-airport static deterministic GHP optimally and heuristics for the single-airport stochastic GHP. In 1991, Richetta [32], in his thesis, developed heuristics for the single-airport dynamic stochastic GHP. In 1993, Terrab and Odoni [41] formulated the single-airport static stochastic GHP with multiperiods as a dynamic programming problem. They proposed several heuristics to solve their version of the GHP and to handle large problem instances that occur in practice. Since the authors were unable to prove that the formulation would yield an integer

solution directly from the linear programming (LP) relaxation, they developed a decomposition method to exploit the fact that the constraint matrix could be partitioned into network matrices. Since this was a static stochastic version of the GHP, the authors described airport capacity in the following manner:

“Capacities are random variables that are given a probabilistic forecast [that can be thought of as] a number of scenarios, each scenario representing a particular instance of the random capacity vector with an associated probability.”

In a ground breaking paper in 1993, Richetta and Odoni [33] used stochastic linear programming with one stage to solve the single-airport static stochastic GHP with multiperiods optimally and expanded previous work by including the dynamic case. They were able to overcome the limitations of the dynamic programming formulation in the paper by Terrab and Odoni [41]. Previous work determined amount of delay to assign on a flight by flight basis. In this paper, the authors formulate the problem by considering flights in the aggregate. This approach allows for the alignment of probabilistic airport capacity with current weather forecasting procedures.

“There are Q alternative scenarios, each scenario providing a possible capacity forecast for the entire time interval of interest with scenario q having a probability equal to p_q .”

At the beginning of the day, specialists at the ATCSCC are given weather forecasts from different sources such as the National Weather Service (NWS), the Aviation Weather Unit, and the meteorologists from different airlines and their AOCs. Each of these forecasts could lend itself to a capacity forecast or scenario.

The decision variables in the model indicate the number of flights of a certain class originally scheduled to arrive at time i , but rescheduled to arrive at time j under a certain capacity scenario q . The inputs are the number of aircraft of a certain class scheduled to arrive at the capacitated airport, the cost of delaying a single flight of a certain class for a given time period on the ground, the cost of delaying a flight in the air, and the number of flights unable to land under a certain capacity scenario.

In 1996, Glockner [18], in his thesis, considered a network of airports and formulated a dynamic time-space network flow problem in which some arc capacities are assumed to be random variables. These arcs are referred to as “restricted arcs” and are used to capture the uncertainty of airport capacity.

In 1999, Hoffman, Ball, Rifkin and Odoni [5] formulated the single-airport static stochastic GHP as an integer programming problem that can be solved in polynomial time. They improve on Richetta and Odoni’s formulation by including fewer decision variables and exploiting the network structure of the problem to find an optimal solution using LP relaxation. As in other stochastic versions of the GHP, arrival capacities are assumed to be random variables:

“Probabilistic information about the uncertain capacity is available in the form of Q scenarios, M_q , for $1 \leq q \leq Q$, where $M_{q,t}$, $1 \leq t \leq T$, is the arrival capacity of the airport during time t , if scenario q is realized. The probability of the q^{th} scenario occurring is p_q .”

The most important contribution of the model in this paper is consistency with the paradigm and procedures of CDM. The inputs to the model are the aggregate demand for each time period, the cost of delaying a single flight on the ground for a single time period, the cost of one period of airborne delay and

the capacity scenarios and associated probabilities. The model minimizes the total expected delay costs, which is the sum of the assigned ground delay and expected airborne delay. The model outputs the optimal number of flights to be delayed on the ground or the number of arrival slots (to make available) per unit time. CDM procedures would then be used to determine the specific flights that should be given ground delay.

The trend is towards a formulation of the GHP that is stochastic in nature because it is a better representation of true conditions during a GDP. Thus, there is a need for the estimation of arrival capacity scenarios with associated probabilities. Methods for determining these arrival capacity scenarios and associated probabilities are the subject matter of this thesis.

2.3 Integer Programming

A linear program (LP) is a constrained optimization problem of the form

$$\begin{aligned} & \min cx \\ & \text{subject to} \\ & Ax \geq b \\ & x \geq 0 \end{aligned}$$

where cx is a linear function, A is an m by n matrix, b is an m -dimensional column vector, and x is an n -dimensional column vector of variables or unknowns. LPs have a long history and many highly efficient commercial codes for solving them are available. An integer program (IP) is an LP with the additional restriction that the x vector can only take on integer values. In particular, the general IP is NP-Hard whereas the LP problem can be solved in polynomial time. The

added restriction of integrality makes the problem more difficult to solve. IP solution methods are very often based on relaxing the integrality constraint and yield the linear programming relaxation of the IP. The optimal solution to the LP relaxation is used as a bound to the optimal solution to the IP. If the IP is a maximization problem, then the optimal value for the LP relaxation is greater than or equal to the optimal value for the IP. On the other hand, if the IP is a minimization problem, then the optimal value for the LP relaxation is less than or equal to the optimal value for the IP. If the LP relaxation is infeasible, then so is the IP. If the optimal solution to the LP relaxation is an integer solution, then this solution is feasible and optimal for the IP. This latter condition is a reflection of the strength of the formulation of the IP. The strongest possible formulation is one in which the solution to the LP relaxation yields an integer solution. This is always the case when A is a totally unimodular matrix.

2.3.1 Totally Unimodular and Interval Matrices

It is desired to have the strongest possible formulation of the IP and that is one in which the optimal solution to the LP relaxation is integer. This is the case when matrix A is totally unimodular (TU). A TU matrix is defined as a matrix in which the determinant of every square submatrix is $+1$, -1 , or 0 . A TU matrix is closed under matrix operations, i.e. the transpose of a TU matrix is TU; a TU matrix with a row or column deleted is TU; a TU matrix that has two rows or columns interchanged is TU; a TU matrix with rows or columns duplicated is TU. There exist matrices with columns that contain more than two nonzero elements, but that still yield a TU matrix. One such matrix is known as an interval matrix, which has the “consecutive ones” property. An interval matrix

is an m by n (0,1) matrix that satisfy the following condition: if $a_{ij} = a_{kj} = 1$ and $k > i + 1$, then $a_{lj} = 1$ for all l with $i < l < k$ [28]. A matrix has the “consecutive ones” property if in any given column, ones appear consecutively. An interval matrix can be transformed into a network matrix using elementary row operations that result in a matrix with network flow properties. If matrix A has the “consecutive ones” property, then the IP can be solved using the LP.

2.3.2 Set Covering and Set Partitioning

Binary IPs, that is, IPs in which the variables are binary, constitute a very important broad class of IPs. Binary variables are used to model problems where a choice has to be made between an event occurring or not occurring. Thus the variables take on the value 1 if the event occurs and 0 otherwise. Set covering and set partitioning problems are special classes of binary IPs.

Let $M = \{1, \dots, m\}$ be a finite set and $\{M_j\}$ a given collection of subsets of M . Let A be the m by n incidence matrix of the family $\{M_j\}$ for $j \in N = \{1, \dots, n\}$. So for $i \in M$,

$$a_{ij} = 1 \text{ if } i \in M_j$$

$$= 0 \text{ if } i \notin M_j ,$$

$$x_j = 1 \text{ if } j \in F$$

$$= 0 \text{ if } j \notin F.$$

We say that $F \subseteq N$ covers M if $\bigcup_{j \in F} M_j = M$. Thus, the set covering problem has the following formulation:

$$\text{Minimize } \sum_{j=1}^n c_j x_j \text{ (} c_j \text{ is the cost of including } M_j \text{ in the cover)}$$

$$\text{subject to } \sum_{j=1}^n a_{ij}x_j \geq 1 \text{ for } i = 1, \dots, m$$

$$x_j \in B^n.$$

The set partitioning problem is formulated in a similar manner with the exception of the constraint being an equality constraint. In general, set covering and set partitioning problems are NP Hard. In the case of problems whose matrices have the consecutive ones property, the IPs can be solved efficiently using the LPs, as stated in the previous section. The columns of the matrices can be enumerated in polynomial time since the number of enumerations is no more than $\binom{n}{2}$, where n is the number of feasible possibilities. A set covering problem can be converted into a minimum cost shortest path problem and solved using standard network flow procedures. In Chapter 5, the problem of assigning months to seasons in a least costly fashion will be formulated as a set covering/partitioning problem with an added constraint of limiting the total number of seasons (covers) chosen. The formulation with this added constraint yields a 0-1 incidence matrix that is not totally unimodular. The matrix has what is known as the consecutive ones property with wrap around. Though the incidence matrix is not TU, a simple iterative procedure can be implemented to determine the optimal solution efficiently.

2.4 Statistical Background

2.4.1 Probability and Empirical Distribution Functions

The probability distribution function (pdf), $f(x)$, of a random variable X is a real-valued function that satisfies the following properties:

- (i) $f(x) > 0, x \in R$;
- (ii) $\sum_{x \in R} f(x) = 1$ for discrete X ,
 $\int_R f(x) dx = 1$ for continuous X ;
- (iii) The probability of event $X \in A$ is:

For X discrete: $P(X \in A) = \sum_{x \in A} f(x)$, where $A \subset R$,

For X continuous: $P(X \in A) = \int_A f(x) dx$.

The function, $F(x)$, defined by

$$F(x) = P(X \leq x) \text{ for } x \in R$$

is known as the cumulative distribution function (cdf) of the random variable X .

The properties of $F(x)$ are:

- (a) $0 \leq F(x) \leq 1$
- (b) If $x' < x''$, then $F(x') \leq F(x'')$.
- (c) $F(\infty) = 1$ and $F(-\infty) = 0$.
- (d) If X is a discrete random variable, then $F(x)$ is a step function and the height of each step at x equals $P(X = x)$.

Thus, cdfs are nondecreasing and right-continuous functions that have a minimum value of zero and a maximum value of one. It is possible to approximate an unknown cdf using an empirical distribution function (EDF), $F_n(x)$, which is completely determined by observed values of a random variable,

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n I\{x_j \leq x\} , j = 1, \dots, n,$$

for each real number x .

An EDF is close to the true cdf for large n and possesses the same properties as a cdf since:

$F_n(x)$ is a relative frequency of $\{x_j \leq x\}$ implies $0 \leq F_n(x) \leq 1$;

$F_n(x)$ is a nondecreasing function since the number of observed values less than or equal to x does not decrease as x increases;

For all values of x less than the smallest observed value, $F_n(x) = 0$;

For all values of x greater than or equal to the largest observed value, $F_n(x) = 1$.

Fitting (probability) distributions to data

Empirical distribution functions are used to estimate underlying cdfs and similarly, there are methods to estimate underlying pdfs using empirical or observed data. The most notable are graphical methods used to summarize the distribution of a univariate data set. Graphical methods allow for the viewing of the location of the observed data, the spread of the data, the skewness of the data, the presence of outliers and the presence of multiple modes. These methods include stem-and-leaf plots, boxplots, bar graphs, probability plots, and relative frequency histograms (See [23] or [36]). The latter form will be the focus in this dissertation. A frequency histogram is constructed by partitioning the range of observed values (largest value minus smallest value) into k equal length subintervals (bases of rectangles or bins) and by calculating the frequency (counts) of observations in each subinterval (heights of rectangles or bins). To construct a relative frequency histogram, simply divide the frequency in bin i , f_i , by the total number of data points, n . A relative frequency histogram estimates an under-

lying pdf because $\sum_i f_i/n = 1$. If a resulting histogram contain many peaks and valleys, a triangular (weighted) binning technique could be used to “smooth” out the histogram. In a smoothed histogram, the frequency in a given bin i is the sum of half of the original frequency in bin i , one quarter of the frequency in the bin on the immediate left of bin i , and one quarter of the frequency in the bin on the immediate right of bin i . Thus, a smoothed histogram can be determined using the following formula:

$$\frac{1}{4}f_{i-1} + \frac{1}{2}f_i + \frac{1}{4}f_{i+1}, \text{ for the total number of bins } k.$$

It should be noted that this is one alternative of choosing the weights for smoothing the histograms. The weights can be chosen in any way that seems appropriate.

Kolmogorov-Smirnov (KS) Goodness of Fit Test

Nonparametric methods (tests) are used when the underlying distribution function is unknown or the assumptions about the underlying population are questionable. These tests include the Wilcoxon Mann-Whitney Test, the Wilcoxon Signed Ranks Test, the Sign Test, the Runs Test and the Kolmogorov-Smirnov Test (See [36]). The KS statistic is used in Section 5.3.4 of this dissertation.

To test whether an EDF comes from a specified formal cdf or to test if two or more EDFs can be assumed to come from the same EDF, a Kolmogorov-Smirnov Test can be used. This test makes no assumptions about the underlying distribution beyond continuity of F , so the KS statistic is referred to as a distribution-free statistic. The KS statistic

$$\text{KS} = \max_x \sqrt{\sum_i \left(\frac{n_i}{n}\right) [F_i(x) - F(x)]^2}, \text{ where } x = 1, 2, \dots, n,$$

measures the maximum deviation of the EDF within classes i to the pooled EDF, F , defined as

$$F = \frac{1}{n} \sum_i (n_i F_i).$$

2.4.2 Time Series

A time series is a sequence of observations ordered in time that are analyzed to ascertain any type of historical pattern, which can be used to create a forecasting model. The goal of analyzing a time series may be any of the following:

- (i) description of the main properties and structure of the time series,
- (ii) explanation of variation of one time series using variation of another,
- (iii) prediction of future values of the time series,
- (iv) control of the process whose quality is measured by the time series.

The most widely used mechanism to view and describe the main properties of a time series is a time plot. Time plots graphically reveal any trends, cycles, seasonal components or irregular fluctuations in a time series. A trend is defined as a long-term upward or downward movement in a time series, a cycle refers to a recurring up and down movement around trend levels, a seasonal component or seasonality refers to periodic movement in a time series that is dependent on the time of year, and irregular fluctuations are erratic fluctuations with no consistent or regular pattern. Smoothing techniques are used to remove random fluctuation and give a clearer view of the behavior of a time series. Smoothing techniques include moving average smoothing and exponential weighted smoothing. The

techniques differ in that past observations are weighted equally for moving averages and unequally for exponential weighted averages. A moving average for time period t , MA_t , is calculated according to the formula

$$MA_t = \frac{(y_t + y_{t+1} + \dots y_{t-N+1})}{N}.$$

In exponential smoothing, exponentially decreasing weights are assigned to observations. More recent observations are weighted more heavily than more remote observations and all observations are referred to as “dampened”. The process of assigning weights starts with the assignment of an initial smoothed value:

$$S_0 = y_1 \text{ (or average of first few observed values of series)}$$

$$S_1 = \alpha y_1 + (1 - \alpha)S_0$$

$$S_2 = \alpha y_2 + (1 - \alpha)S_1$$

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \text{ where } \alpha \text{ is a smoothing constant}$$

$$\text{and } 0 < \alpha \leq 1, t \geq 1.$$

The series S_t is the smoothed estimate for time period t and equals a fraction of the newly observed time series value for time period t plus a fraction of the smoothed estimate made in time period $t-1$. An appropriate value of the smoothing constant, α , can usually be determined from historical data. The basic equation for single exponential smoothing is

$$S_t = \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i y_{t-i} + (1 - \alpha)^t S_0.$$

Exponential smoothing can be thought of as “geometric” smoothing since the weights $\alpha(1 - \alpha)^t$ decrease geometrically with time t , for $0 < \alpha \leq 1$. Moving average smoothing techniques will be used to perform an analysis of seasonality in Section 5.3.1. For a more in-depth treatment of time series analysis, see Chatfield [10].

2.4.3 Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is used to study the effects of one or more independent (predictor) variables on the dependent (response) variable. Most commonly, ANOVA is used to test the equality of means by analyzing the total sum of squares (about the combined mean), which is partitioned into different components (due to model or due to random error).

Single-Factor or One-Way

A factor is an independent variable. Thus in single-factor ANOVA, the effects of only one independent variable are being tested. For single-factor ANOVA, each level of the factor is referred to as a “treatment”. The null hypothesis is equality of factor level means. The Single-Factor ANOVA model can be written as

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

where Y_{ij} represents the i^{th} observation of the j^{th} factor level

$$i = 1, \dots, n_j, j = 1, \dots, k,$$

n_j is the number of observations for the j^{th} factor level, k is the total number of factor levels, μ is the overall mean of all factor level means, and α_j is called the effect of the j^{th} factor level.

The unknown parameters (μ, α_j) are usually estimated from the data using the method of ordinary least squares (OLS). In OLS, $\sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \mu - \alpha_j)^2$ is minimized with respect to $\mu, \alpha_1, \alpha_2, \dots, \alpha_k$. As stated earlier, the equality of factor level means are tested by analyzing the decomposition of overall variance (total sum of squares). The deviation $(Y_{ij} - \bar{Y}_{..})$, the difference between each observation and the overall mean can be decomposed into two components: deviation between each factor level mean and the overall mean; and the deviation of each observation around its respective factor level mean. Because

$$Y_{ij} - \bar{Y}_{..} = \bar{Y}_{.j} - \bar{Y}_{..} + Y_{ij} - \bar{Y}_{.j}$$

where:

$$\bar{Y}_{.j} = \frac{\sum_i Y_{ij}}{n_j}$$

$$\bar{Y}_{..} = \frac{\sum_j \sum_i Y_{ij}}{n}, \quad n = \sum_{j=1}^k n_j,$$

we have the identity

$$\sum_j \sum_i (Y_{ij} - \bar{Y}_{..})^2 = \sum_j n_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 + \sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})^2$$

after squaring and then summing the deviations. In words, we say that the “total sum of squares,” equals the sum of the “sum of squares due to model (treatment sum of squares)” plus the “sum of squares due to random error (error sum of squares).” Each sum of squares term divided by its associated degrees of freedom results in its mean square (MS). The F -value that is used as a test statistic is the ratio of the mean square of the model and the mean square error. Mean square of the model can also be thought of as the mean squared deviation between groups (treatments) and the mean squared error as the mean squared deviation within groups. Large values of the F -statistic lead to the rejection of the null hypothesis

Source of Variation	SS	df	MS	F
Between Groups	$SS_{Model} = \sum_j n_j (\bar{Y}_{.j} - \bar{Y}_{..})^2$	$k - 1$	$\frac{SS_{Model}}{k-1}$	$F =$
Error(Within Groups)	$SS_{Error} = \sum_j \sum_i (Y_{ij} - \bar{Y}_{.j})^2$	$n - k$	$\frac{SS_{Error}}{n-k}$	$\frac{MS_{Model}}{MSE}$
Total	$SS_{Total} = \sum_j \sum_i (Y_{ij} - \bar{Y}_{..})^2$	$n - 1$		

Table 2.2: ANOVA Table for One-Way Layout

of the factor level means being equal. All aforementioned values are summarized in an ANOVA table (Table 2.2), which any statistical software will output.

Residual Analysis

The residual term, $\hat{\varepsilon}_{ij}$, in the Single-Factor ANOVA model is the difference between an observed value and its fitted (estimated) value:

$$\hat{\varepsilon}_{ij} = Y_{ij} - (\hat{\mu} + \hat{\alpha}_j) = Y_{ij} - \bar{Y}_{.j}.$$

In order for the results of an F -Test to be valid, there are assumptions on the error terms that must be met: they must be independent, have zero mean, constant variance (known as homoscedasticity), and must follow a normal distribution. The adequacy of an ANOVA model is checked by graphical residual analysis. An overall plot of the residuals should appear normally distributed with mean zero. If the observations are from a time series, then the residuals could be plotted in time order. One can say that the assumptions are satisfied if the time plot of residuals appear to be a “horizontal band” and free of any systematic behavior. If the plot appears funnel-shaped, then this is an indication that the constant variance assumption has been violated. Serial correlation or “autocorrelation” between successive residuals would violate the independence assumption. Serial

correlation can be identified by plotting $\hat{\varepsilon}_t$ versus $\hat{\varepsilon}_{t-1}$ and looking for linear trends. The Durbin-Watson Statistic [29],

$$\frac{\sum_{t=1}^{N-1} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^N \hat{\varepsilon}_t^2},$$

may be used to test for serial correlation between successive residuals. This statistic is calculated by dividing the sum of squared differences in successive residuals by the sum of squared residuals. The value of the statistic ranges between 0 and 4. A value close to 0 indicates positive autocorrelation whereas a value close to 4 indicates negative autocorrelation.

There are many methods that could be implemented if the assumptions on the error terms are not satisfied. If the errors do not have homogeneity of variance, then transformations can be made on the original data to stabilize the variance. These transformations include log or square root transformations (used when variance is a function of the mean). An alternative method is applying a weighted least squares method on the data. Weighted least squares can be used instead of transformations if error variances are known completely. WLS can improve the accuracy of estimates even if variances are not completely known, but testing of hypotheses cannot be done without knowing the error variances. The weights in the analysis are used to adjust the amount of influence of observations with larger error variances on the model parameters.

Multiple Comparisons

When the F -test rejects the null hypothesis that there exists an equality of means, the procedure of multiple comparisons allows one to determine where the differ-

ences lie while controlling the simultaneous confidence coefficient $(1-\alpha)$. In general, the procedure of multiple comparisons is used to determine if there exists statistically significant differences between two or more factor level means. Each comparison is known as a contrast, L , and is defined as

$$L = \sum t_j \mu_j$$

where t_j satisfies the restriction $\sum t_j = 0$.

There are three common methods of multiple comparisons that are used: the Tukey Method, the Scheffe' Method and the Bonferroni Method. The Tukey method should be used when the factor level sample sizes are equal and the multiple comparisons of interest are all pairwise comparisons. Scheffe's method is the most general method in that it can be used regardless of whether or not the factor level sample sizes are equal and when all possible comparisons are sought. The Bonferroni method can be used irregardless of the factor level sample sizes, but only for a prespecified set of contrasts. The method that yields the greatest amount of precision of the confidence intervals depend on the type and amount of multiple comparisons being made. See [29] for more details on single-factor ANOVA and multiple comparison procedures.

Chapter 3

Description of Data: Sources and Preparation

The statistical models in this thesis are formulated based on data received from the Federal Aviation Administration (FAA) and the National Climatic Data Center (NCDC). The reliability of the models is strongly dependent on the amount and quality of data that is used for model formulation and calibration.

In the following chapters, models will be presented for the estimation of a vector of capacity scenarios or airport arrival capacity distributions. These models are formulated using ground delay program parameters and weather conditions at San Francisco International Airport (SFO). The data collected are from two main databases:

- Ground Delay Program (GDP) Logs in which data is recorded at the ATC-SCC and archived by Metron, Inc.
- Surface Airways Hourly (TD3280) Weather Data, which is collected and archived by the NCDC.

Section 3.1 describes the type of parameters that are found in the GDP database and provides summary statistics of the data. Section 3.2 defines the

Variables	Definition of Variables
Airport	airport at which GDP is run
Date	date of the institution of GDP
Start Time	planned start time of GDP
End Time	planned end time of GDP
CNX Time	cancellation time of GDP
AAR	airport acceptance rate (set by ATCSCC)
Center(s)	centers included in scope of GDP
Max Delay	maximum delay incurred during GDP

Table 3.1: Data from GDP Logs

variables contained in the weather database while Section 3.3 explains the difficulties and sources of error when trying to merge the two databases. Finally, Section 3.4 describes the data contained in Aggregate Demand Lists (ADLs) that is used in the decision models of Chapter 6.

3.1 Ground Delay Program (GDP) Logs

On a daily basis, the specialists at the ATCSCC record all facility operations from the beginning of the day until the end of the day. These records are included in a daily report known as the Command Center Logs. One main component of these logs are records of the national GDPs, the procedures and parameters at which they were run and a digest and critique of the actions taken during a GDP. The GDP logs are taken directly from GDP cover sheets that contain an hourly record of all parameters and weather conditions pertinent to a certain GDP. Table 3.1 gives a list and description of the variables that are found in the GDP Logs.

month:	1	2	3	4	5	6	7	8	9	10	11	12
1995	28	13	17	14	20	13	15	6	15	13	16	24
1996	20	22	9	7	10	12	9	8	18	11	16	22
1997	24	9	7	14	3	5	8	15	7	9	11	7

Table 3.2: Number of GDPs at SFO

It should be noted that there were some minor differences between the logs from year to year. For example, in 1995, only a “duration” was given instead of a “start time” and “end time”. It is a trivial matter to extract these fields from the duration. There is also evidence of continuous improvements to the database through the addition of more fields. One of the difficulties in working with real data was a lack of consistency in the archived format.

Basic statistics were calculated on the GDP logs for SFO. Such statistics include number of GDPs per month, percentage of GDPs canceled, percentage of GDPs planned during morning hours and the actual average duration of the morning GDPs. These statistics will play a vital role in the formulation of the models which are presented in later chapters.

According to Table 3.2, there is no apparent trend in the numbers of GDPs. What is of more interest is the percentage of GDPs planned and implemented that were canceled. No less than 25% of the programs planned at SFO are canceled for any given month in any given year (See Figure 3.1). This is an alarming statistic which confirms the need for better predictive models to improve GDP effectiveness.

Figure 3.2 gives the percentage of GDPs that are planned and implemented during the morning hours, 1600Z-1900Z (8 am-11am local time). More than 50%

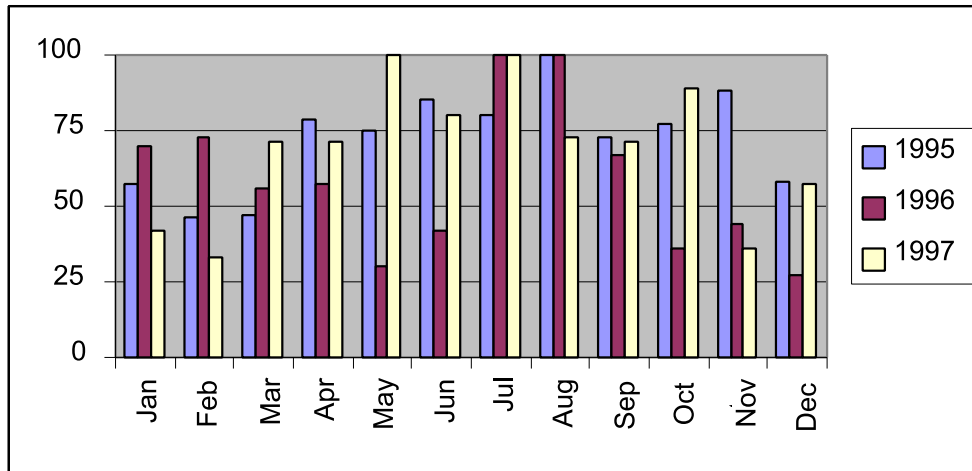


Figure 3.1: Percentage of GDPs Canceled at SFO

of all programs are morning GDPs at SFO. This is principally due to the onset of fog in the morning during the peak demand hours. Due to this high probability, all analysis in this thesis are performed on morning GDPs at SFO.

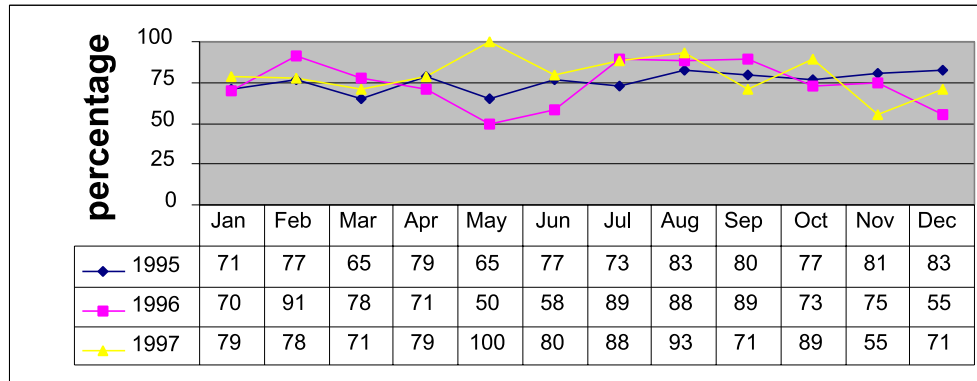


Figure 3.2: Percentage of GDPs Implemented During Morning Hours at SFO

3.2 Surface Airways Hourly Weather Data

Weather conditions and runway configurations play the principal role in determining airport capacities. Recall Table 2.1 and Figure 2.1. Since it is our goal to predict and estimate arrival capacity distributions, it is imperative that we consider weather conditions. Some of the models in this thesis are calibrated using GDP parameters and weather data is used to support the assumption that the use of distributions of GDP durations are appropriate surrogates for distributions of IFR conditions (in the case of morning fog at SFO) during instances when weather data is not available. This was shown by comparing burnoff times of fog at SFO with cancellation/end times of GDPs. Burnoff times for marine stratus conditions during summer months were obtained from MIT's Lincoln Laboratory. It was determined that a GDP's cancellation time was within 25 minutes from the burnoff time of fog conditions during the months May-September.

Distributions of IFR conditions at SFO were derived using data from "Surface Airways Hourly" from the NCDC during the years 1984-1992. This database contains an hourly listing of surface weather observations that are taken at stations located primarily at major airports and military bases. See Table 3.3 for a listing and description of the data fields. These stations are operated by the National Weather Service (NWS), the U.S. Air Force Weather Service, the U.S. Navy Weather Detachment and the Federal Aviation Administration (FAA). Data at these stations are collected using the Automated Surface Observing System (ASOS) that was designed primarily for aviation operations. The weather variables used to determine distributions of IFR conditions are ceiling height and visibility. Recall that IFR conditions are marked by a ceiling below 1000 feet or a visibility less than 3 miles.

Variables	Description
Year	(self-explanatory)
Month	(self-explanatory)
Day	(self-explanatory)
Hour	(self-explanatory)
CLHGT	height of cloud ceiling in hundreds of feet
VISIB	visibility in miles
WEATHER	coded weather conditions (Table 3.4)
TEMP	temperature in whole degrees Fahrenheit
DWPT	dew point temperature in whole degrees Fahrenheit
WIND (direction)	wind direction in whole degrees
WIND (speed)	velocity of wind in knots

Table 3.3: Data from Surface Airways Hourly

Weather Codes
F: Fog
H: Haze
K: Smoke
L: Drizzle
Z: Freezing
W: Shower
R: Rain
S: Snow
IP: Ice Pellets

Table 3.4: Codes for Weather Data in Surface Airways Hourly

3.3 Merging GDP Logs and Weather Data

There are many difficulties when trying to merge two databases. One must determine the conditions on which to match in order to merge the databases. The biggest drawback is the possible loss of information due to the lack of similar fields or data. One possible cause for the difficulty is the fact that the range of dates for weather data is 1984 to 1992 whereas the range for the GDP data is 1995 to 1997. Thus, there is no common data on which to merge the databases.

From a statistical point of view, the sample size of 3 years worth of data is relatively small and inadequate to make definitive statistical conclusions. It is well-known in the aviation and aviation weather community that reduction in arrival capacity is due primarily to low ceiling and limited visibility conditions [49]. In this dissertation, arrival capacity distributions will be solely based on a conditional distribution of duration of IFR conditions given that a GDP is planned. Chapter 4 will give a more detailed explanation about the derivation of this conditional distribution.

3.4 Aggregate Demand Lists

At the start of each day, the AOCs submit their daily flight schedules (arrivals and departures) over an “extra-net” known as the CDMnet. The CDMnet facilitates data exchange between the AOCs and the ATCSCC and works in conjunction with CDM procedures. Dynamic updates to schedules are submitted through the CDMnet and viewed on FSM. The Volpe Transportation Systems Center merges the data from the airlines and other data from the NAS to create aggregate demand lists (ADLs). Dynamic changes such as delays or cancellations of flights

Air	Airline
Flt	Flight ID
OETD	Original Estimated Time of Departure
OETA	Original Estimated Time of Arrival
OETE	Original Estimated Time EnRoute
ETE	Estimated Time EnRoute
LCTD	Last Controlled Time of Departure
LCTA	Last Controlled Time of Arrival
ERTA	Earliest Runway Time of Arrival
ARTD	Actual Runway Time of Departure
ARTA	Actual Runway Time of Arrival

Table 3.5: Data fields in the Aggregate Demand Lists

are incorporated into the ADLs that are sent out via the CDMnet approximately every 5 minutes [27]. Data in the ADLs contain information on a flight-by-flight basis. Table 3.5 lists the flight data fields that are used in the decision models of Chapter 6.

Chapter 4

Airport Arrival Capacity Distributions (Capacity Scenarios)

A capacity scenario is one possible sample path or realization of capacity over a given interval of time, thus each scenario can be thought of a time series. The capacity scenarios derived in this thesis are the required inputs into the Hoffman-Rifkin Static Stochastic Ground Holding Model, as well as other stochastic ground holding models. In this dissertation, each capacity scenario will be referred to as an arrival capacity distribution (ACD). Since each ACD is one possible realization of capacity, we desire to derive a vector of possible ACDs with associated probabilities.

4.1 Types of Arrival Capacity Distributions

On any day, there is a given weather forecast that translates into a particular capacity. As the weather (forecast) changes, so does the capacity. The severity of the weather and the accuracy of the forecast determine the amount of fluctuation in the capacity level. It is a normal practice for specialists at the ATCSCC

to receive different weather forecasts from various sources. Each of these forecasts could realistically result in a different capacity scenario or arrival capacity distribution (ACD). The strength of the forecast could possibly determine the probability of a particular ACD.

We now describe, on a conceptual level, a range of possible ACD models. Just as is the case for the Hoffman-Rifkin model, we assume that time is discretized and the ACD models are based on this assumption. In any ACD, the x-axis represents time of day, y-axis represents arrival capacity levels or number of flights able to land, and each bar represents arrival capacity over a given time interval, usually in one hour or 15-minute increments.

In general, an ACD can take almost any structure imaginable. Thus the number of possible ACDs is enormous. In this section, the four most representative ACDs for almost any airport will be described. In the most general ACD model, there can be a constant fluctuation in the arrival capacity level, as seen in Figure 4.1. We shall refer to this as the “general” ACD because it can be used to model almost any given airport that may be plagued with constant weather or runway configuration changes. In order to estimate this most general ACD model, each arrival capacity level (level of reduced capacity), start time of reduced capacity and end time of reduced capacity must all be estimated.

A simpler model allows for only 2 capacity levels and the distribution fluctuates between these 2 levels. This type of model may adequately represent conditions at an airport with few runway configurations or with one main weather pattern that have many peaks throughout the day. This will be known as the “2-Level” ACD (Figure 4.2).

A further simplification of the 2-Level ACD models has a structure in which

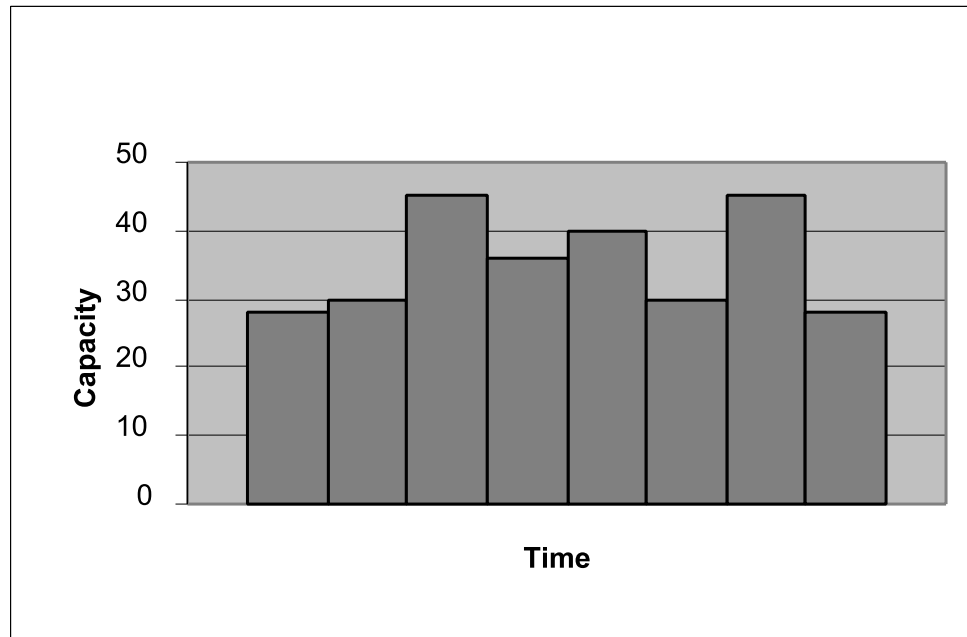


Figure 4.1: Form of the General Arrival Capacity Distribution

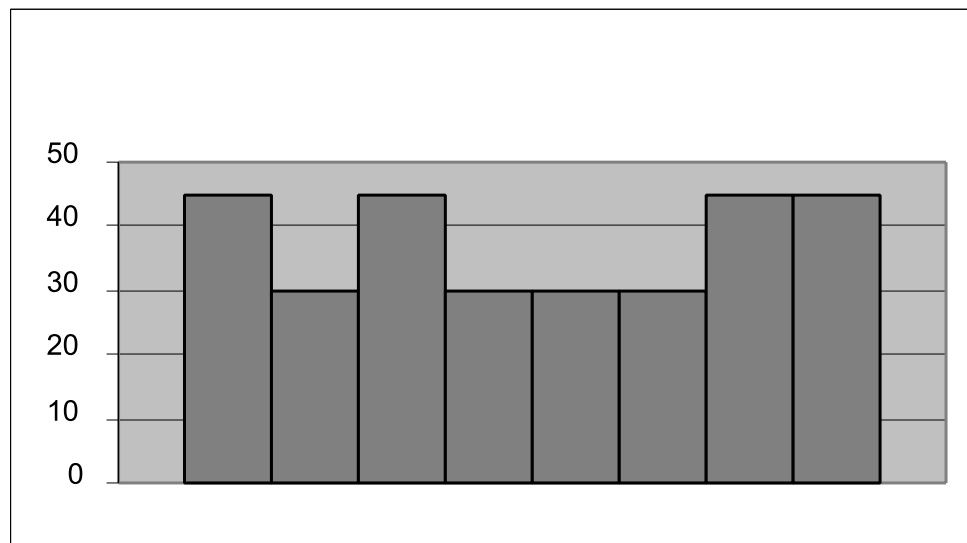


Figure 4.2: Form of the 2-Level Arrival Capacity Distribution

there is “normal” or maximum capacity at the beginning of the day, then reduced capacity for a certain length of time, followed by a return to the “normal” level. This model may capture most airports in which there is a consistent weather pattern that occurs sometime after sunrise and only lasts for a finite length of time. This ACD is appropriate for most airports whose arrival capacity level is cut almost in half when inclement weather forces a change from VFR to IFR approaches. In this case, we need only estimate the two parameters: start time and duration of reduced capacity. Thus, this is referred to as the “2-Parameter” ACD (Figure 4.3).

The simplest ACD is a distribution that has reduced capacity at sunrise, remains constant at this level for a given time and then increases to the normal arrival capacity level. Therefore, the only parameter to be estimated is the end time (duration) of the reduced capacity. Hence, this ACD model is the “1-Parameter” ACD (Figure 4.4). This is a reasonable way to model ACDs associated with weather patterns that are present at sunrise, remain continuously in place for a period of time and then clears.

In general, the task of estimating a vector of ACDs is quite daunting because of the possible model complexity of an individual ACD. There is a range of model complexities, which can be seen in the ACD models presented here. It was discovered that the structure of the two simplest models, the 2-Parameter and 1-Parameter ACDs, are representative of actual capacity scenarios for a reasonably broad range of airports. In fact, the 1-Parameter ACD can be applied in one very important and practical case. It can be used to model morning fog at San Francisco’s International Airport (SFO). This case of modeling morning fog at SFO is the focus of this dissertation.

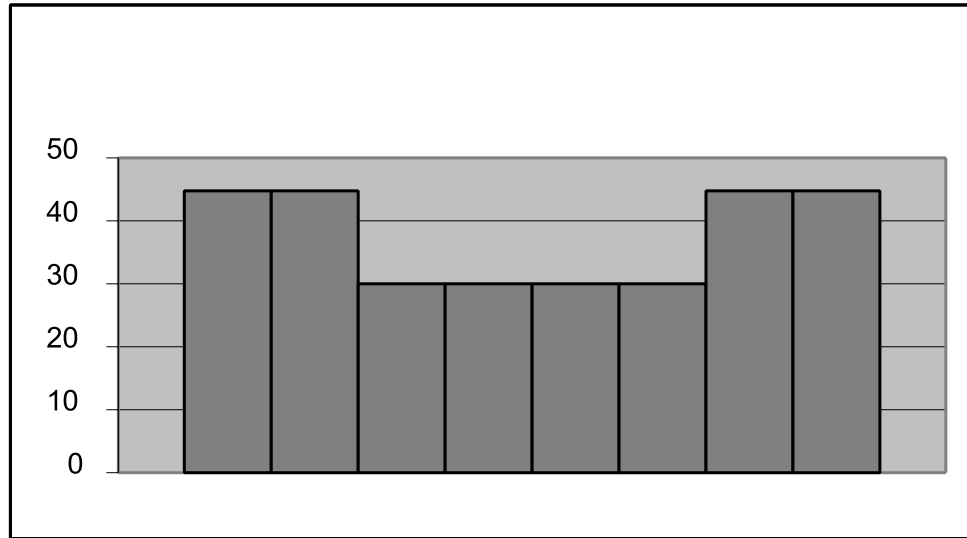


Figure 4.3: Form of 2-Parameter Arrival Capacity Distribution

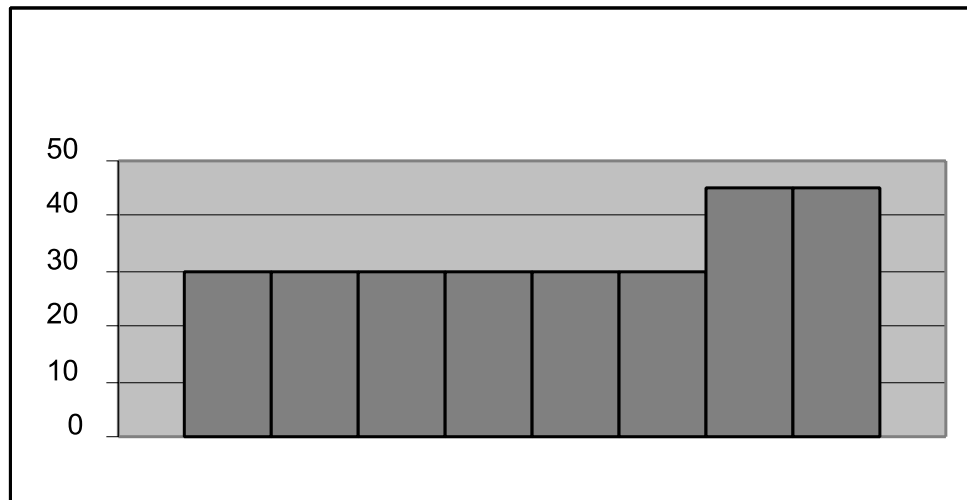


Figure 4.4: Form of 1-Parameter Arrival Capacity Distribution

4.2 Estimating 1-Parameter ACDs at SFO

In the general case, we would have to estimate the reduced capacity value (IFR capacity level), the start time of reduced capacity and the end time of reduced capacity in order to estimate a general ACD. We will initially consider the case of early morning GDPs that are typically caused by fog. A good example of where this occurs is SFO. At SFO, morning fog exists at sunrise and lasts for a given (stochastic) period of time. When the fog burns off, the arrival capacity level returns to the “normal” value of 45 flights per hour (when visual approaches can be performed, See Table 2.1). For this situation, we are only required to estimate the duration of reduced capacity, since the start time is assumed to be at sunrise and the IFR capacity for SFO is 30 flights per hour (See Figure B.1 in appendix).

Nominal IFR conditions at SFO are characterized by a ceiling of less than 2500 feet or a visibility of less than 3 miles. Since fog conditions are present at sunrise, the start of reduced capacity is at sunrise. To estimate the duration of reduced capacity, the duration of a GDP will be calculated. Morning GDPs at SFO are planned to end at the burnoff time of fog or at the initial dissipation of the stratus conditions.¹ Recall that in Chapter 3, results were given to suggest that the end time or cancellation time of a GDP corresponds to the burnoff time of marine stratus conditions at SFO during the summer months. According to [11], stratus clouds form during overnight hours and dissipate during the morning hours. There are times when the dissipation or burnoff of the “cloudiness” occurs after the late morning arrival traffic peak (1800Z) in which demand is high. During these times, there exists a capacity-demand imbalance due to fog. Therefore, restricting attention to the morning GDPs at SFO is reasonable for capturing the

¹From an e-mail correspondence with Forrest Terral at the ATCSCC on June 15, 1999.

length of time of IFR conditions due to fog during peak demand times. Since fog materializes during times of low demand, weather data can be used to give an estimate of the duration of IFR conditions irregardless of demand.

GDP durations on days when GDPs were planned and durations of nominal IFR conditions can be used for estimates of the durations of reduced capacity in 1-Parameter ACDs. Thus, there are two sources of data that can be used to estimate 1-Parameter ACDs for SFO. Since the Hoffman-Rifkin static, stochastic ground holding model requires a vector of ACDs, the vector will be determined by estimating a distribution of 1-Parameter ACDs.

4.3 Determining Distribution of 1-Parameter ACDs

In general, a distribution can be determined by binning a given set of observations or empirical data to create a relative frequency histogram (see Section 2.4.1 for more detail). The 3 years of empirical GDP data from SFO are used to create the frequency histogram of duration of morning GDPs (Figure 4.5). This histogram will be determined by considering the duration of GDPs conditioned on a GDP being planned.

Notice the peak between 4 and 5 hours in Figure 4.5. This is due to the operational procedures at the ATCSCC during the given 3 years of data. There was a limit on the maximum number of hours a GDP could be run and that limit was 4 hours. To remove the effect of operational procedures on the overall distribution, a smoothing technique described in Section 2.4.1 is utilized. To smooth out peaks and valleys in the histogram, a triangular binning technique is used. In this technique, half of the number of data points in a given bin is added to one-quarter the number of data points in the bin on the (immediate) left side

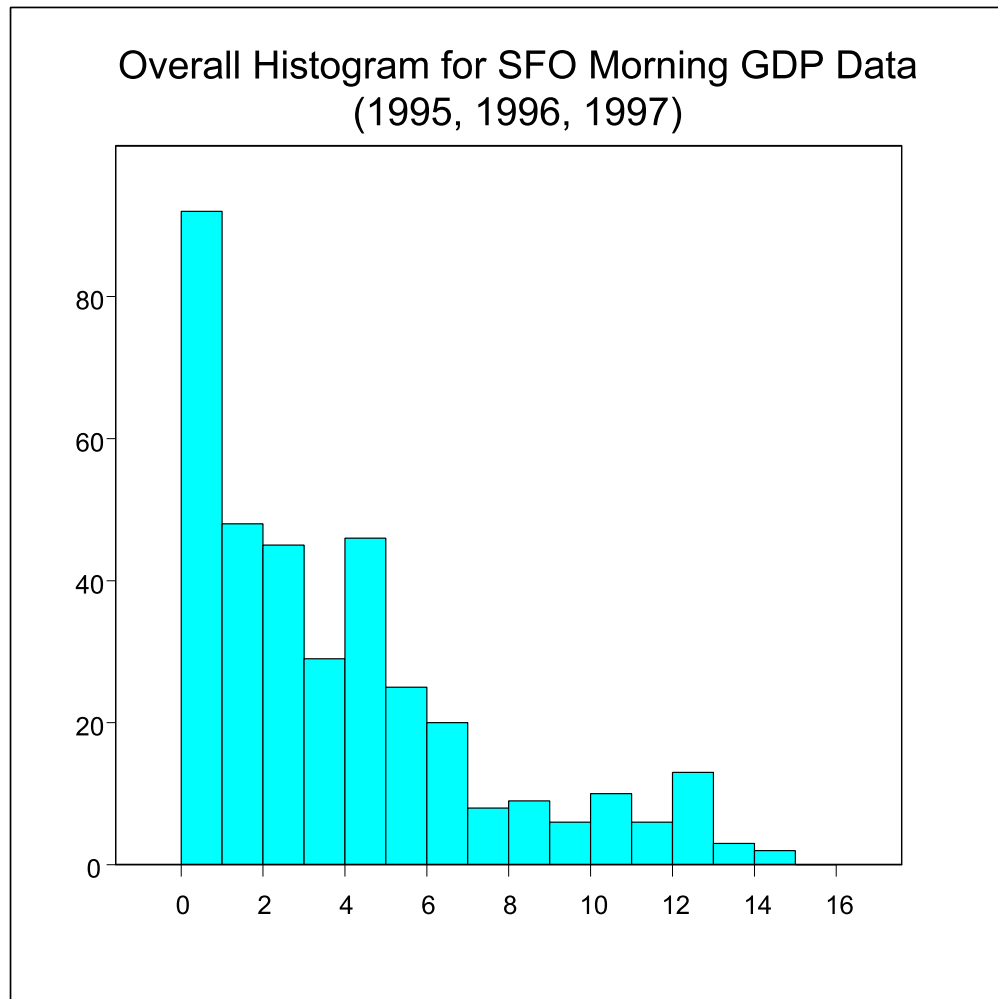


Figure 4.5: Overall Frequency Histogram for SFO Morning GDP Data

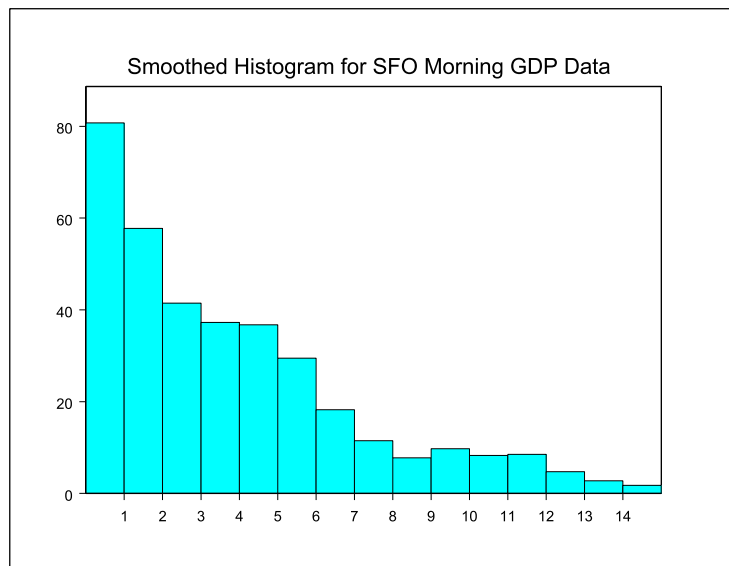


Figure 4.6: Smoothed Histogram for SFO Morning GDP Data

of the given bin and one-quarter the number in the bin on its right side to give the new frequency or number of data points in the given bin. This process is repeated for all bins except the first and last bins. To account for end effects, smoothing of “end” bins is done by taking the weighted sum of three-quarters of the frequency in the given bin plus one-quarter of the frequency in the bin immediately next to the given bin. The smoothing of all bins result in a “smoothed frequency histogram” (Figure 4.6).

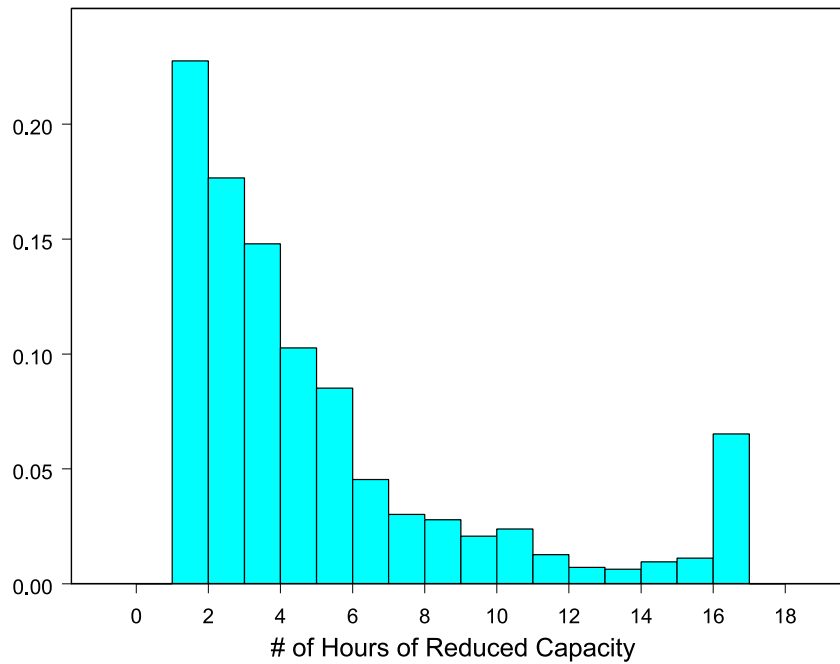


Figure 4.7: Relative Frequency Histogram for Duration of IFR Conditions

Each bin of a histogram corresponds to the duration of reduced capacity in a particular ACD. Thus the histogram can be thought of as the distribution of ACDs. To get an associated probability of an ACD (or bin on histogram), simply divide the frequency in a given bin by the total sum of frequencies (relative frequency):

$$\mathcal{P}(S_i) = \frac{\text{Frequency}_i}{\text{Sum of Frequencies}}, \text{ for each scenario (ACD) } S \text{ and bin } i.$$

It can be argued that a distribution based on weather data is more representative of durations of IFR conditions. Figure 4.7 gives this type of distribution using weather data with the relative frequencies along the y-axis.

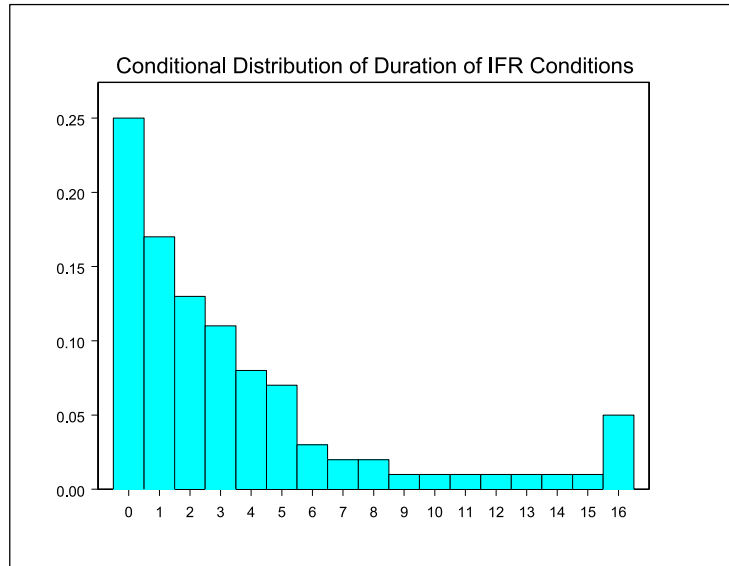


Figure 4.8: “Conditional” Distribution of Durations of IFR Conditions

This distribution is a conditional distribution because it is the distribution of duration of IFR conditions given that the duration equals or exceeds one hour. Thus, there is no zero bin in this distribution, but there is a zero bin on the GDP distribution. Since the goal is to estimate the duration of IFR conditions during

instances of capacity-demand imbalances for which a GDP will be implemented, some reasonable combination of the two distributions is sought. One alternative is to derive a conditional distribution of the durations of IFR conditions given that a GDP is planned. Recall that the distribution of GDP durations is a conditional distribution: distribution of GDP durations given that a GDP is planned. If the 0-bin on the GDP data distribution is referred to as $P(GL = 0|\text{GDP is Planned})$, then the “conditional” distribution of duration of IFR conditions given that a GDP is planned is a weighted combination of $P(GL = 0|\text{GDP is Planned})$ and $P(\text{IFR Durations}|\text{IFR} \geq 1)$. To determine this “conditional” distribution, we simply include the zero bin from the GDP distribution in the distribution of IFR conditions and normalize accordingly to derive the new associated probabilities. See Figure 4.8.

Histograms are used to give information about an underlying pdf of empirical data. In this section, histograms created using ALL available empirical (GDP and weather) data were presented. The underlying pdf will be referred to as a Capacity Probabilistic Distribution Function (CPDF) and is the vector of 1-Parameter ACDs that will be used as input into the Hoffman-Rifkin model. The underlying CPDF is a distribution that is based on weather conditions that are highly stochastic in nature. It is feasible to think that the CPDF would change according to the changes in weather. Thus, the fundamental mechanism that controls the CPDF is continuously changing over time. A CPDF can be created from any given set of observations, in this case, for the set of years of available historical data at SFO. There are different types of CPDFs that can be derived by partitioning the overall CPDF into subunits based on the underlying changing mechanism (weather). In the next chapter, models are presented that are the

result of partitioning the overall CPDF in different ways.

Chapter 5

Determining Capacity Probabilistic Distribution Functions (CPDFs) that Vary in Time

There is a class of stochastic ground holding models that are solved using probabilistic distributions of “scenarios” or possible realizations of arrival capacity as input (along with other inputs). These models are described in [2], [5], [18], [21], [33], [34], and [35]. ACDs and the distribution of ACDs, CPDFs, could be used in any of the models of this class. In this chapter, methods for partitioning (overall) CPDFs in different ways will be presented. Chapter 6 will discuss the Hoffman-Rifkin static stochastic model and how to input the results of this chapter into that decision model.

Given empirical data about capacity (or IFR conditions), relative frequency histograms can be constructed and used to estimate CPDFs. Depending on how the underlying weather mechanism (forecast) changes, daily, monthly or seasonal CPDFs are all types of distributions that can be estimated. The type that will ultimately be utilized depends on the operational preferences of the specialists at the ATCSCC, as well as other factors.

5.1 Daily CPDFs

Each day of a year is subject to a particular long-term climatic condition so it may be desired to create daily histograms (CPDFs). In the case of GDPs, with only 3 years of available data, each daily histogram would only contain 3 data points. This is not a very informative histogram, nor much of a histogram at all. There are nine years of weather data on which to create a daily histogram, but the sample of nine points is still too small to be able to estimate an underlying CPDF for the daily histogram. To account for this problem of sample size, the data for a particular day is augmented by data from preceding and succeeding days. One alternative for creating a histogram for a given day is by considering data 15 days prior to the given day, 15 days after the given day and the given day itself. This is reasonable because weather conditions are usually similar in a given month (31 days).

There are many ways to weight the data during the 31 days used in a given day's histogram. The first alternative is to just weight each day's duration equally since it is assumed that the weather is more or less consistent during these 31 days. As an example, suppose the histogram for January 1st needs to be derived using the 3 years of available GDP data. Then, the length on January 1st, the lengths for the 15 days prior to January 1st and the lengths for the 15 days after January 1st are used to create the January 1st histogram. The histograms (Figures 5.1- 5.3), created using GDP data in the manner described above, reflect the assumption of consistent weather.

Another way to weight the data is by a triangular weighting technique. In this case, one gives more weight to the data during the week that contains the day for which the histogram is being created (e.g. .5), less weight to the days

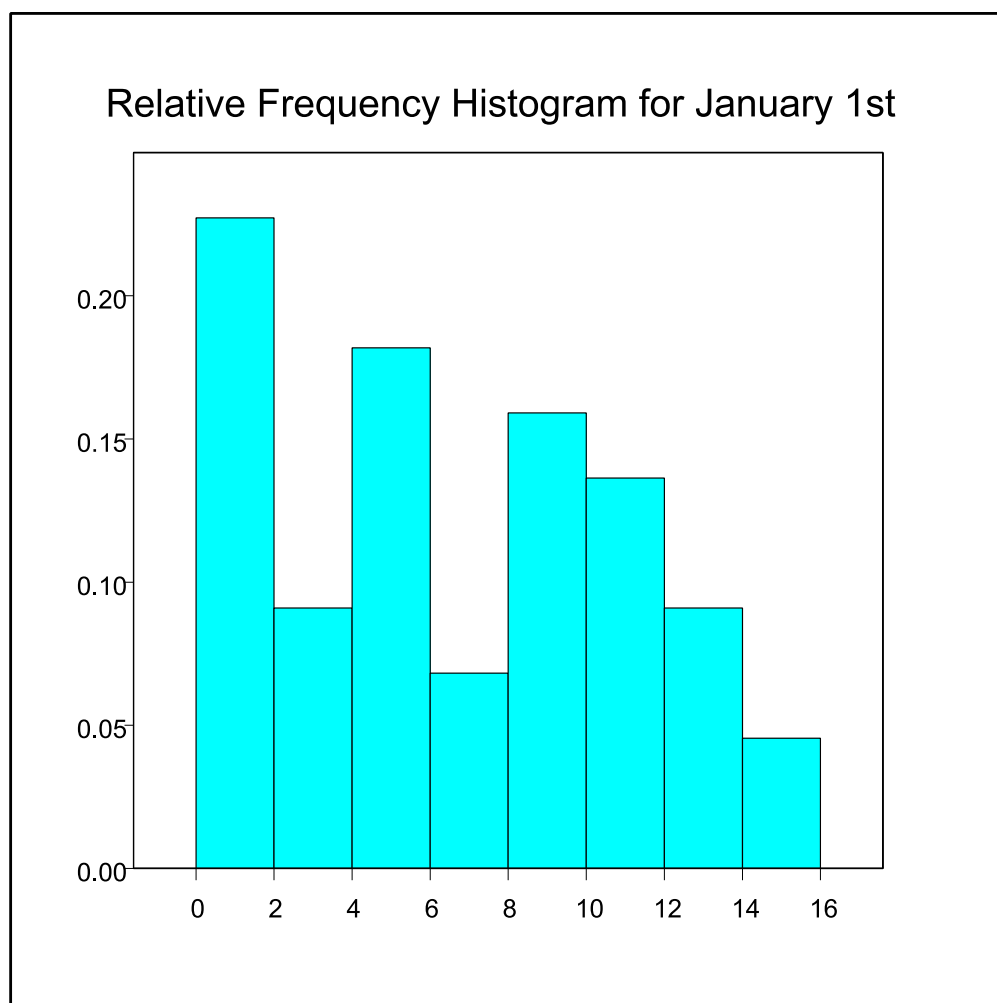


Figure 5.1: Relative Frequency Histogram (Uniform Weighting) for January 1st

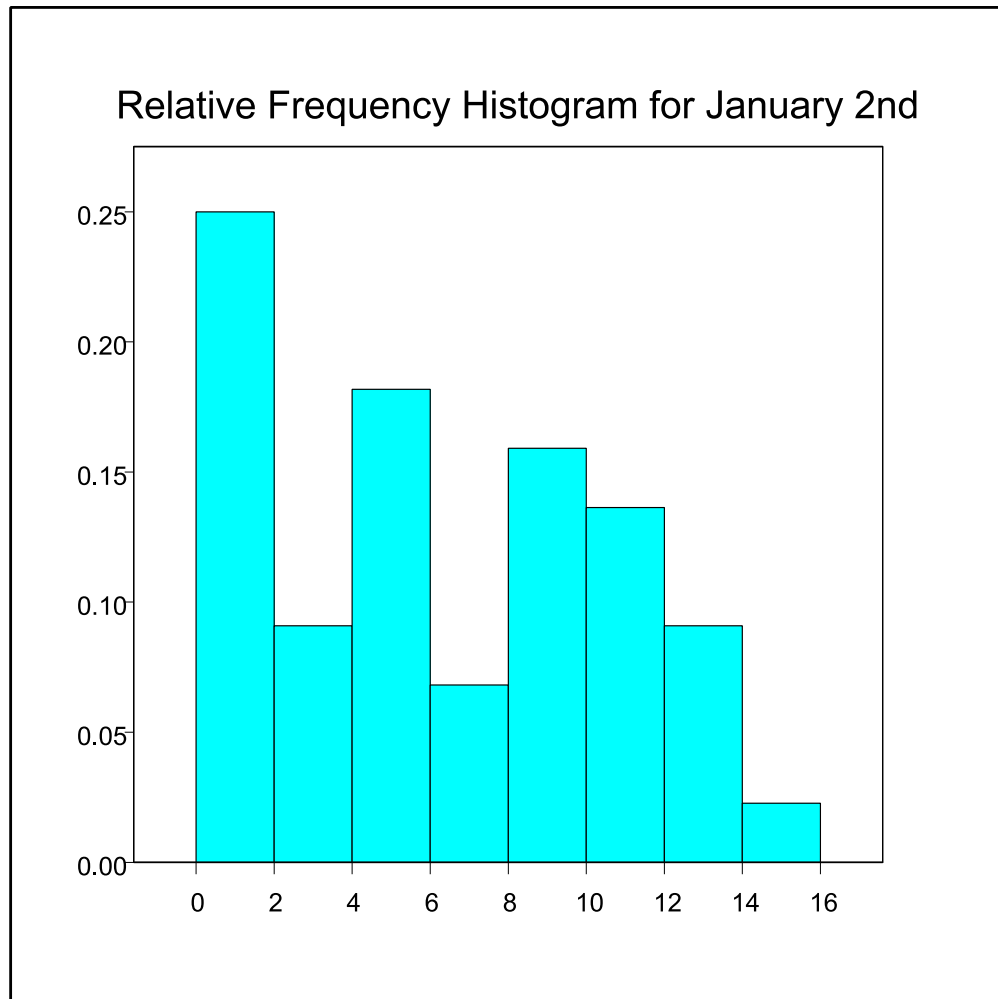


Figure 5.2: Relative Frequency Histogram (Uniform Weighting) for January 2nd

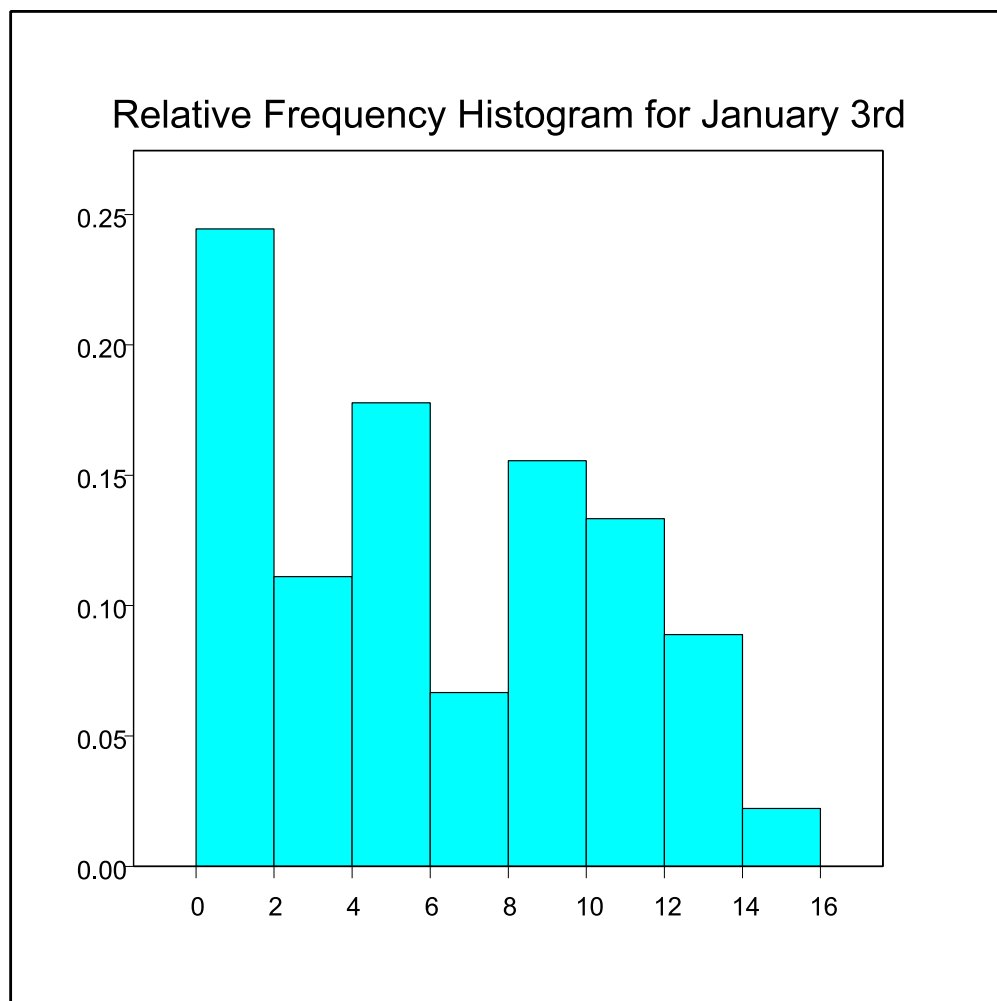


Figure 5.3: Relative Frequency Histogram (Uniform Weighting) for January 3rd

contained in the second week centered around the given day (e.g. .3) and the least weight to the data in the third week centered at the given day (e.g. .2). There are various weighting techniques that could be utilized. It ultimately depends on the amount of confidence one has in the consistency of the underlying changing mechanism for a group of days.

5.2 Monthly CPDFs

Another way to address the problem of sample size for daily distributions is to group the daily data by month. This grouping would yield monthly CPDFs. It is possible that several months may have the same or similar CPDFs, especially in the case of distributions of IFR or inclement weather conditions. Weather conditions such as thunderstorms and snow occur at certain times of year or during specific seasons. At SFO, the most prevalent weather conditions are radiation fog and advection fog. Radiation fog is also known as ground fog and occurs when the temperature drops to the dew point near the ground. Advection fog occurs when warm, moist air moves over a colder land mass. According to the Weather Sensing Group at MIT's Lincoln Laboratory, radiation fog occurs more than 100 days annually and advection fog is the next most frequently occurring weather condition at SFO. (See Figure B.2) Based on conversations with specialists at the ATCSCC, fog is heaviest from September to the middle of March and burnoff times are difficult to ascertain. Through the Marine Stratus Initiative at SFO [11], which is led by the Weather Sensing Group at Lincoln Laboratory, it has been determined that the stratus cloud season is during the months of May to September. As an example, a possible CPDF may be the same during the months September-March and the same during the months May-September, but

different for the two groupings of months. The next section will present methods for determining seasonal CPDFs. It can be assumed that seasonal GDPs correspond to seasons of certain weather conditions. Thus, monthly distributions will be grouped into seasons, based on some measure of similarity, to create seasonal CPDFs.

5.3 Seasonal CPDFs

Decomposing an overall CPDF into groupings of months (seasons) based on some measure of similarity (dissimilarity) is reminiscent of partitional clustering in which data is partitioned into disjoint clusters. This partitioning is done by minimizing a measure of dissimilarity within each cluster and maximizing the dissimilarity between different clusters. The data used for this dissertation are time ordered, so a simple clustering technique is not adequate. A method is needed to perform clustering that is imbedded in a time series. This type of clustering will be referred to as seasonal clustering. The resulting clusters must be contiguous and homogeneity should exist within the clusters. Since the data constitute a time series, time plots will be used to detect seasonal trends as an initial step in determining seasonal clusters.

5.3.1 Detecting Seasonal Trends

In Section 2.4.2, the possible ways of detecting seasonal trends were discussed. The one used in this thesis is the method of plotting time series data in a time plot. This is an effective technique to gauge an overall trend in time-ordered data.

Figure 5.4 is a simple time series plot of average GDP lengths from month to month for 1995 through 1997. Observe that the months October to March lie above the horizontal line (average GDP length = 3 hours) and the months May to September lie below the line. This observation gives credence to the idea that the CPDFs would differ between these two groupings of months, as stated in the previous section.

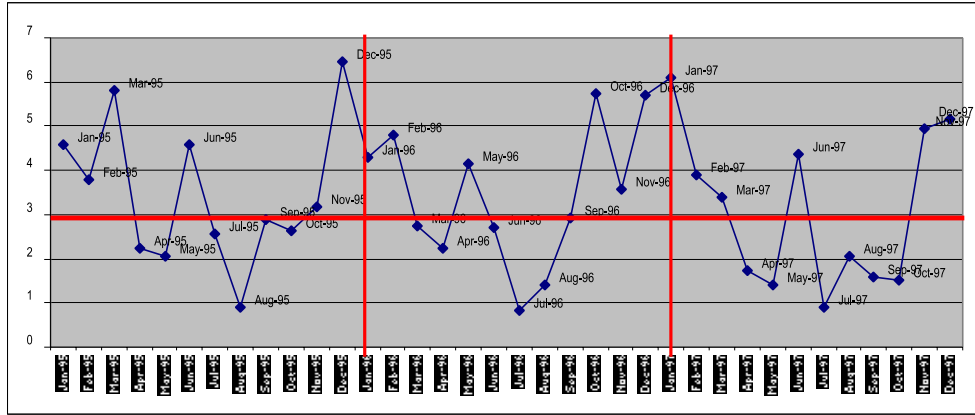


Figure 5.4: Time Series Plot of Average GDP Durations

There appears to be a seasonal trend in the time series plot of monthly average GDP Durations (Figure 5.4). A plot can also be created from the actual daily GDP lengths. To remove random fluctuation, a moving average smoothing technique is implemented. A 15-day centered moving average (CMA) is calculated for a given day using its GDP length (GL) (or length of IFR conditions) and the lengths of GDPs (IFR conditions) 7 days prior to the desired day and 7 days after the desired day:

$$CMA_t = \frac{GL_{t-7} + \dots + GL_t + \dots + GL_{t+7}}{15}.$$

Time series plots of averages of 15-Day CMAs of GDP durations and durations of IFR conditions are Figures 5.5 and 5.6, respectively. The general trend appears

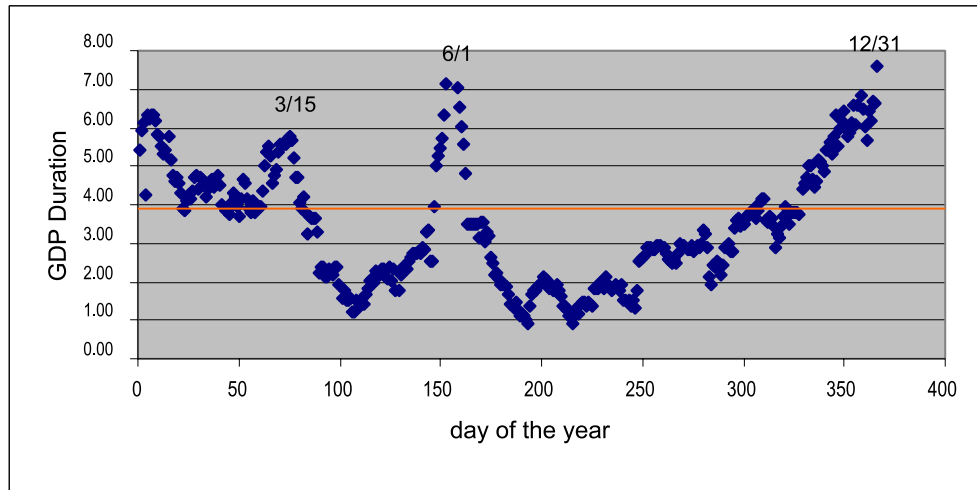


Figure 5.5: Average Centered Moving Averages of GDP Durations

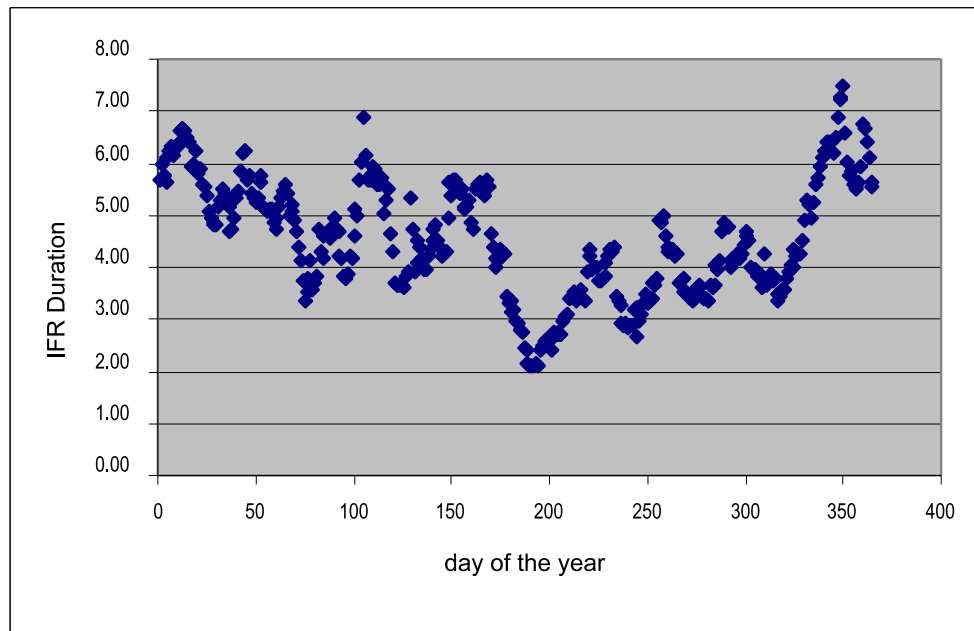


Figure 5.6: Average Centered Moving Averages of Durations of IFR Conditions

to be similar in the time plots, but there are some differences. These differences can be attributed to demand. In order for a GDP to be planned, the demand for an airport's arrival resource must be expected to exceed the available capacity. Thus, a GDP is in effect as long as demand is high. Thus, the hours for a GDP, as represented in Figure 5.5, are only during peak demand hours. The durations of IFR conditions are not dependent on demand and thus, may be longer than the durations of GDPs. This observation is reflected in Figure 5.6. In the next sections, model formulation and solutions of a more rigorous analysis to ascertain seasonal clusters are presented.

5.3.2 Model Formulation

Given the twelve months in a year, the goal is to partition the year into groupings of (contiguous) months that contain the most similar weather conditions. The problem of determining the optimal partitions (seasons) can be formulated as a minimum cost shortest path problem or as a set covering/partitioning integer programming problem. We note that the shortest path approach has a restriction associated with it. It can not handle the case where a season starts in one month and ends in a month at the beginning of the year. This is what is known as “wrap around.” We will note later how this shortest path model can be imbedded into an iterative procedure that can handle this more general case.

In general, a shortest path problem is a network optimization problem that seeks to find the smallest distance between a given (source) node, node 0, and all other nodes. Consider a network G with m nodes, n arcs and a cost c_{ij} associated with each arc (i, j) in the network G . The length (cost) of a (directed) path is the sum of the lengths of the arcs in the path. The shortest path problem can be

thought of as a network in which it is desired to send one unit of flow at minimal cost. The shortest path problem can be formulated as an integer program:

$$\begin{aligned}
& \text{Minimize} && \sum_{i=0}^m \sum_{j=0}^m c_{ij} x_{ij} \\
& \text{Subject to} && \sum_{j=0}^m x_{ij} - \sum_{j=0}^m x_{ji} = \begin{cases} 1 & \text{if } i = 0 \\ 0 & \text{if } i \neq 0 \text{ or } m \\ -1 & \text{if } i = m \end{cases} \\
& && x_{ij} = 0 \text{ or } 1 \quad i, j = 0, 1, \dots, m.
\end{aligned}$$

Here, node 0 is the origin (source) node of the shortest path and node m is the destination. The constraints of the shortest path problem are known as flow conservation equations and yield a node-arc incidence matrix that is totally unimodular, hence its optimal solution is the optimal solution to the LP relaxation.

In the case of partitioning a year into groupings of contiguous months, let m equal the number of elements to be clustered ($m = 12$ months) and let $D = [d_{i,j}]$ be a dissimilarity matrix, where $d_{i,j}$ is the cost of grouping contiguous months i to j . The clustering problem is defined on an acyclic network with $m+1$ nodes as illustrated in Figure 5.7. In this network, an arc from node i to node j represents forming the cluster starting at element $i+1$ and ending at element j . The cost of arc (i, j) , c_{ij} , is equal to $d_{i+1,j}$.

According to Bodin [9], the shortest path from node 0 to node m “in a graph that contains exactly k branches defines a clustering of the original data into k categories.” This shortest path is determined by minimizing the sum of dissimilarities along the arcs of the path. (In [37], Saigal gives a dynamic programming algorithm to solve this such problem). Hence, the problem of determining the

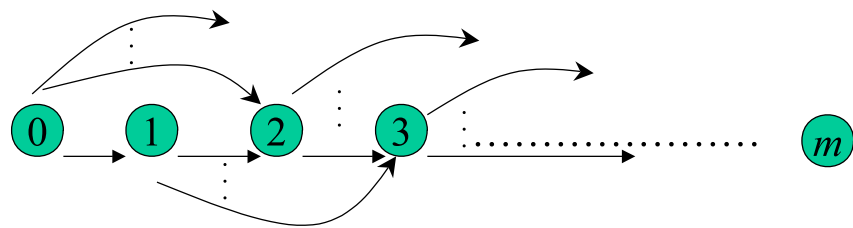


Figure 5.7: Network Representation of Seasonal Clustering Problem

groupings of months (seasons) of the year into k seasons can be solved by finding the shortest path from node 0 to node m containing exactly k arcs.

The problem can also be solved by a set covering/partitioning integer programming model. The goal of the set covering integer program is to “cover” the whole year by a finite number of covers or seasons with the smallest total cost and the goal of the set partitioning IP is to cover the whole year by a finite DISJOINT set of seasons in a least cost fashion. Section 2.3 contains a discussion of integer programming and set covering. Recall that the set covering/partitioning IP has the following formulation:

$$\begin{aligned} & \text{Minimize} && \sum_{j=1}^n C_j x_j \\ & \text{subject to} && \sum_{j=1}^n a_{ij} x_j \geq 1, \text{ for each month } i \\ & && x_j \in \{0,1\} \end{aligned}$$

and the formulation for the set partitioning IP is:

$$\begin{aligned} & \text{Minimize} && \sum_{j=1}^n C_j x_j \\ & \text{subject to} && \sum_{j=1}^n a_{ij} x_j = 1, \text{ for each month } i \\ & && x_j \in \{0,1\} \end{aligned}$$

where $A=[a_{ij}]$ is a 0-1 incidence matrix with $a_{ij} = 1$ if $i \in M_j$ (month i is in candidate season M_j), 0 otherwise; $\{M_j\}$ correspond to candidate seasons, $j = 1, \dots, n$; n is the number of candidate seasons; C_j is the cost of including M_j in the cover; and x_j is a binary variable with value 1 when M_j is included in the cover, and 0 otherwise.

In this case of assigning months to seasons, the columns of A (i.e. the set of M_j 's) can be efficiently enumerated since a season is characterized by a start month and an end month and the months must be contiguous. The possible seasons can be enumerated according to length of (contiguous) months. If all possible combinations are allowed, i.e. groupings 1 month in length (M_1, \dots, M_{12}), 2 months in length (M_{13}, \dots, M_{24}) up to 12 months in length, there are a total of 133 possible seasons. Since there are 12 months, there are 12 different seasons for each possible season length except for the season of length 12 (only 1 way). Thus, 12 multiplied to 11 plus 1 results in 133 possible seasons. Intuitively, no weather season lasts more than 5 months. If the length of the season is restricted to being no more than 5 months, then there is a total of 60 possible seasons. Since there are 12 months and 5 different possible season lengths, enumerating the seasons yields 60 possible seasons in this candidate season set. Results will be given for both the candidate season sets. The incidence matrix A (for a candidate season set of size 60) is:

	M_1	M_2	...	M_{12}	M_{13}	M_{14}	...	M_{24}	...	M_{49}	M_{50}	...	M_{60}
$A =$	<i>Jan</i>	1	0		0	1	0		1		1	0	1
	<i>Feb</i>	0	1		0	1	1		0		1	1	0
	<i>Mar</i>	0	0		0	0	1		0		1	1	0
	<i>Apr</i>	0	0		0	0	0		0		1	1	0
	<i>May</i>	0	0		0	0	0		0		1	1	0
	<i>Jun</i>	0	0		0	0	0		0		0	1	0
	<i>Jul</i>	0	0		0	0	0		0		0	0	0
	<i>Aug</i>	0	0		0	0	0		0		0	0	0
	<i>Sep</i>	0	0		0	0	0		0		0	0	1
	<i>Oct</i>	0	0		0	0	0		0		0	0	1
	<i>Nov</i>	0	0		0	0	0		0		0	0	1
	<i>Dec</i>	0	0		1	0	0		1		0	0	1

Observe that the 0-1 incidence matrix, A , almost has the consecutive ones property (Section 2.3.1); for example, the ones in columns M_{13} , M_{14} , M_{49} and M_{50} are consecutive. Recall that a matrix having this property is totally unimodular (TU), and thus, the IP can be solved as an LP. This is a desired property because LPs can be efficiently solved using commercial software whereas IPs are, in general, more difficult. The consecutive ones property does not hold because there are wrap around columns such as M_{24} and M_{60} in matrix A . Though a matrix with consecutive ones and wrap around is not TU, it can be solved in polynomial time using a simple iterative procedure. This procedure involves rotating rows to delete the wrap around column and solving the LP for each rotation. The solution chosen is the best solution of all the optimal solutions of the rotations.

We note that a similar iterative procedure could be applied to make use of the shortest path model to solve set partitioning problems with wrap around.

Partitioning the months of the year into a finite number of seasons containing contiguous months can be formulated as either a minimum cost shortest path problem or a set covering/partitioning integer programming problem. The formulation chosen for this thesis is the set covering/partitioning formulation and it will be used in subsequent sections.

5.3.3 Determining Seasonal Clusters using Average GDP Durations

In this section, a seasonal “clustering” technique that assigns consecutive months to a particular season based on some measure of similarity will be developed. A way to derive a finite number of seasons that contain contiguous months in a least costly fashion is desired. A set covering/partitioning integer program model will be used to determine the seasons.

Since the seasons are chosen in a least costly fashion, a cost of a season must be defined and determined. Conceptually, the cost of season M_j , C_j , is the “difference” between a month’s CPDF and a season’s CPDF. In this analysis, the cost function will be based on a difference in means. While this clearly represents an approximation, it should be noted that it appears that there exists a direct (increasing) relationship between the mean and the variance of GDP Lengths (Figure 5.8). Hence, a cost function based on comparing means should also capture differences in variances, in this case.

Several cost functions are possible for comparing seasonal and individual monthly means. In this section, different cost functions will be given and com-

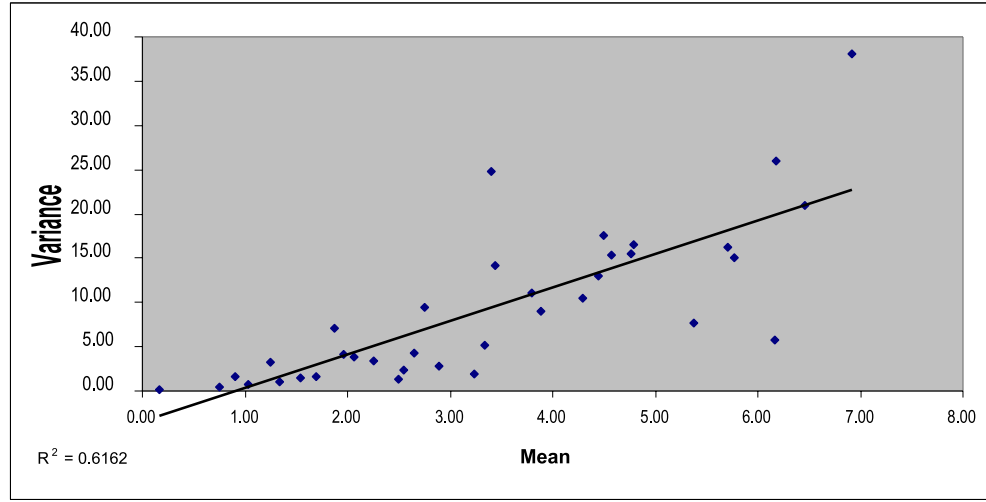


Figure 5.8: Relationship Between Mean and Variance of GDP Duration

pared. Section 5.3.5 will discuss a way to evaluate the quality of a given set of seasons based on a broader set of criteria.

The following cost functions are considered: (i) sum of squared deviations (SoSqs), (ii) normalized sum of squared deviations, or (iii) seasonal variances. The cost functions were chosen because they measure the difference between a season's mean and the means of the months contained in the season. The first cost is the sum of squared deviations between a season's value (average GDP duration) and the values of the months contained within that season. The cost, normalized sum of squared deviations, is the sum of squared deviations divided by the number of months contained in that season. This cost function is chosen due to the possibility of a longer season being penalized by having a larger value for SoSqs. A seasonal variance is deemed appropriate because actual daily ground delay durations are considered. A seasonal variance is determined by calculating the variance of all daily GDP durations from the overall seasonal average. Table 5.1 gives the formulas for the three different clustering criteria. Here $\bar{X}_{.j}$ is the

Sum of Squared Deviations(SoSqs)	$\sum_{j=1}^m (\overline{X}_{.j} - \overline{X}_{..})^2$
Normalized SoSqs	$(\frac{1}{m}) \sum_{j=1}^m (\overline{X}_{.j} - \overline{X}_{..})^2$
Seasonal Variances	$(\frac{1}{m-1}) \sum_j \sum_i (X_{ij} - \overline{X}_{..})^2$

Table 5.1: Seasonal Clustering Criteria (Cost Functions)

average over all days i in month j , $\overline{X}_{..}$ is the (overall) seasonal average over all days i and all months j , and X_{ij} is the GDP length on day i in month j .

As an example, the cost of the January/February GDP Season using SoSqs is calculated:

$$\begin{aligned}
C_{13} &= (GDPAvg_{13} - GDPAvg_1)^2 + (GDPAvg_{13} - GDPAvg_2)^2 \\
&= (4.62 - 5.08)^2 + (4.62 - 4.16)^2 \\
&= .2116 + .2116 \\
&= .4232
\end{aligned}$$

where 13 denotes the January/February Season, 1 denotes month January and 2 denotes month February.

The candidate set of seasons must be enumerated and input into the set covering/partitioning model. Each season has a value: the average duration of a GDP in that season. For example, the value for January is the average of the Jan95 average GDP duration, Jan96 average GDP duration and Jan97 average GDP duration. The value for the Jan/Feb season is the averages of all GDP average durations for Jan95, Jan96, Jan97, Feb95, Feb96 and Feb97. It is possible that the set covering/partitioning procedure, under certain seasonal clustering criteria, could choose all seasons of length 1. Hence, a constraint

limiting the number of seasons chosen is added to the set covering formulation.

Now our set covering problem (with added constraint) can be formulated as follows:

$$\begin{aligned}
& \text{Minimize} && \sum_{j=1}^n C_j x_j \\
& \text{subject to} && \sum_{j=1}^n x_j \leq N \\
& && \sum_{j=1}^n a_{ij} x_j \geq 1, \text{ for each month } i \\
& && x_j \in \{0,1\}
\end{aligned}$$

where N is the maximum number of covers or seasons.

To solve this set covering problem, the CPLEX Linear Optimizer 6.0 on a SUN Sparc10 Station was used. Table 5.2 gives the set covering solutions in terms of seasons for $n = 60$. Observe for $N=4$, there is “over-covering” that occurs using the seasonal variance cost function. If set partitioning is used, then the resulting seasons (Mar-Jul, Aug, Sep, Oct-Feb) are disjoint. This approach seems more appropriate for a seasonal “clustering” method since the results of the set partitioning model ensures disjoint clusters. This issue is discussed in more detail in Section 5.3.6.

It is interesting to note that for $N=4$ and $n=60$, the seasons determined by the SoSqs’ seasonal clustering criterion correspond to the boxed seasons in the time plot of the monthly average GDP lengths averaged over all 3 years, 1995, 1996 and 1997 (Figure 5.9).

In the remainder of the dissertation, the seasons for this particular result will be referred to as: Winter GDP Season (Nov/Dec/Jan/Feb/Mar), Spring GDP Season (Apr/May/Jun), Summer GDP Season (Jul/Aug), and Fall GDP Season (Sep/Oct). The relative frequency histograms that estimate the CPDFs for these seasons are given in the appendix (Figures C.1, C.2, C.3, C.4). Notice

For 60 Possible Seasons:	N = 3	N = 4	N = 5
SoSqs	<div>Apr-Jun</div> <div>Jul-Oct</div> <div>Nov-Mar</div>	<div>Apr-Jun</div> <div>Jul/Aug</div> <div>Sep/Oct</div> <div>Nov-Mar</div>	<div>Apr-Jun</div> <div>Jul-Sep</div> <div>Oct/Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div>
Normalized SoSqs	<div>Apr-Aug</div> <div>Sep/Oct</div> <div>Nov-Mar</div>	<div>Feb-Jun</div> <div>Jul-Sep</div> <div>Oct/Nov</div> <div>Dec/Jan</div>	<div>Apr/May</div> <div>Jun</div> <div>Jul/Aug</div> <div>Sep/Oct</div> <div>Nov-Mar</div>
Seasonal Variances	<div>Apr-Aug</div> <div>Jul-Oct</div> <div>Nov-Mar</div>	<div>Apr-Aug</div> <div>Jul-Oct</div> <div>Jul</div> <div>Nov-Mar</div>	<div>Mar-Jun</div> <div>Jul</div> <div>Aug</div> <div>Sep</div> <div>Oct-Feb</div>

Table 5.2: Set Covering Solutions of GDP Seasons (n=60)

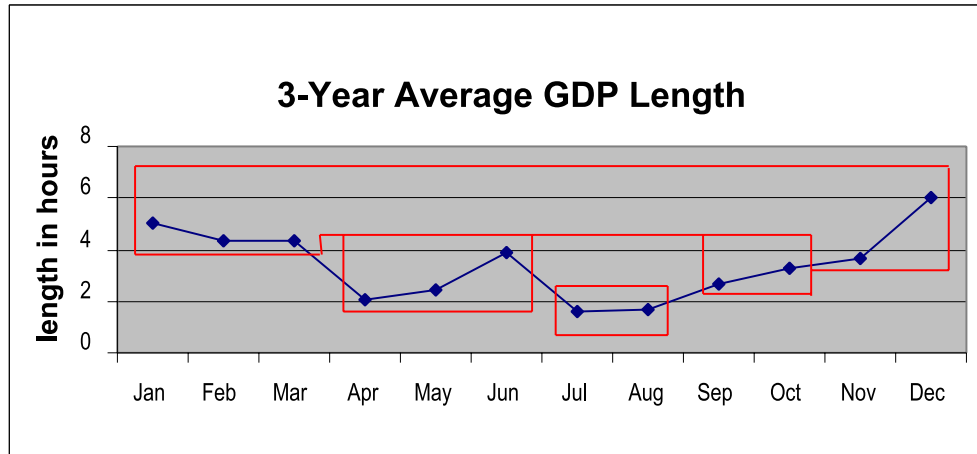


Figure 5.9: Average of Monthly Average GDP Durations over 1995, 1996, 1997

that some of these histograms have a peak at 4 hours and “valleys” at other hours, thus the smoothing technique is implemented and results are given in Figures C.5, C.6, C.7, and C.8. In Table 5.3, the relative frequencies or associated probabilities of ACDs in a particular season are given for both the frequency histograms and the smoothed histograms. (The set covering and set partitioning results for the candidate season set containing 133 seasons are given in Table B.1 in the appendix.)

The results of this section assume that the seasons have fixed monthly boundaries. Since it is possible to have a season that begins or ends on an arbitrary day, a method is needed to determine seasons with arbitrary boundaries. A set covering formulation can also be used in this case. The objective of the set covering IP would be to minimize seasonal costs of all possible seasons with arbitrary start and end days.

In this section, for fixed seasonal boundaries, the overall distribution using GDP data was partitioned into seasonal clusters or subunits. The same proce-

	0	1	2	3	4	5	6	≥ 7
Winter (Nov-Mar)	.21	.09	.07	.06	.15	.08	.08	.26
Winter-Smoothed	.14	.12	.08	.09	.12	.11	.07	.27
Spring (Apr-Jun)	.35	.15	.10	.12	.13	.06	.01	.08
Spring-Smoothed	.23	.21	.13	.13	.12	.07	.04	.07
Summer (Jul/Aug)	.35	.29	.17	.10	.05	.02	0	.02
Summer-Smoothed	.27	.30	.20	.12	.06	.03	.01	.01
Fall (Sep/Oct)	.21	.10	.27	.08	.12	.10	.08	.04
Fall-Smoothed	.14	.18	.19	.15	.11	.11	.08	.04

Table 5.3: ACD Probabilities for Frequency and Smoothed Histograms

dures could be applied using the weather data. In the next section, the clustering criterion developed will be based on differences in distributions rather than differences in means. This criterion will be applied to both the GDP data and the weather data.

5.3.4 Determining Seasonal Clusters using Empirical Distribution Functions

In the previous section, CPDFs were based only on means due to the relationship between the means and variances. Since it is possible to have two distributions that have the same mean, but are different, the cost function (in this section) will be based on differences in distributions instead of differences in means.

Recall from Chapter 2 that an empirical distribution function (EDF) is used to estimate an underlying cumulative distribution function (cdf) of a group of observations or empirical data. The Kolmogorov-Smirnov (KS) test is used to

test if two or more samples come from the same distribution. Since the KS statistic measures the maximum deviation between the EDFs within classes and the pooled EDF, it will be used as the cost of a season in the cost function of the set covering/partitioning formulation.

For any given season in the candidate season sets of size 60 or 133, an EDF is calculated for each month j in the given season according to:

$$F_j(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{I}\{x_k \leq x\}$$

where n is the number of data points (days of GDPs or IFR conditions) in month j . For each real number x , $F_j(x)$ calculates the proportion of data that is less than or equal to that point x . The average of the monthly EDFs, known as the pooled EDF, gives the EDF for the season. The pooled EDF, $F(x)$ is computed by:

$$F(x) = \frac{1}{n} \sum_j n_j F_j$$

where n_j is the sample size for month j and $n = \sum_j n_j$. The KS statistic is appropriate for measuring the difference in a season's EDF and the EDFs of the months contained in that season. The KS statistic will be used as the cost of a given season in any of the candidate season sets and is calculated as:

$$\max_x \sqrt{\sum_j \left(\frac{n_j}{n}\right) [F_j(x) - F(x)]^2}$$

A season whose KS statistic is small implies that the maximum deviation of any month's EDF from the seasonal EDF is small. Hence, the objective of the set covering/partitioning formulation is to minimize the maximum deviation of the months' EDFs from the seasonal EDF or minimize the KS statistic for a given season. Note that the KS statistic requires two or more classes (months) in

order to be calculated, thus no single month seasons are allowed using this cost criterion. Table 5.4 contains the resulting seasons using both the GDP data and the weather data for 60 candidate seasons. Table B.2 in the appendix give the results for the candidate season set of size 133.

The seasonal clustering criterion (cost function) developed in this chapter have yielded many different sets of seasons. A method is needed to assess the quality of the sets of seasons and to determine which set is the “best” set.

5.3.5 Post Analysis for Evaluating Sets of Seasons

In previous sections, different cost functions yielded different sets of seasons. The question now is: which set of seasons is the “best” set? This section will discuss different methods for evaluating the quality of a given set of seasons. One way to evaluate a given set of seasons is by comparing the means of the different seasons to ascertain if they are statistically different from each other. This can be done using the method of multiple comparisons in a single-factor analysis of variance (ANOVA). Single-factor ANOVA is used to test whether there do indeed exist statistically differences in the means of the months. This must be performed before multiple comparisons because if there does not exist a difference in means (null hypothesis not rejected), then there is no need to determine where the differences are.

Section 2.4.3 discussed assumptions that must be met in order for the results of the F -Test of single-factor ANOVA to be valid. These assumptions can be checked using graphical residual analysis. Figure 5.10 demonstrates that the assumption of constant error variance or homoscedasticity is violated since the time plot has a “dumbbell” shape. This implies that the day to day variability in

For 60 Possible Seasons:	N = 3	N = 4	N = 5
Covering (GDP Data)	<div>Apr/May</div> <div>Jun-Oct</div> <div>Nov-Mar</div>	<div>Apr-Jun</div> <div>Jul/Aug</div> <div>Sep-Jan</div> <div>Feb/Mar</div>	<div>Apr/May</div> <div>Jun-Oct</div> <div>Oct/Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div>
Partitioning (GDP Data)	<div>Apr/May</div> <div>Jun-Oct</div> <div>Nov-Mar</div>	<div>Apr-Jun</div> <div>Jul/Aug</div> <div>Sep-Jan</div> <div>Feb/Mar</div>	<div>Apr-Jun</div> <div>Jul/Aug</div> <div>Sep-Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div>
Covering (Weather Data)	<div>Mar-Jun</div> <div>Jul-Sep</div> <div>Oct-Feb</div>	<div>Mar/Apr</div> <div>May/Jun</div> <div>Jul-Sep</div> <div>Oct-Feb</div>	<div>Mar/Apr</div> <div>May/Jun</div> <div>Jul-Sep</div> <div>Oct/Nov</div> <div>Nov-Feb</div>
Partitioning (Weather Data)	<div>Mar-Jun</div> <div>Jul-Sep</div> <div>Oct-Feb</div>	<div>Mar/Apr</div> <div>May/Jun</div> <div>Jul-Sep</div> <div>Oct-Feb</div>	<div>Mar/Apr</div> <div>May/Jun</div> <div>Jul-Sep</div> <div>Oct/Nov</div> <div>Dec-Feb</div>

Table 5.4: Set Covering/Partitioning Solutions using KS statistics (n=60)

GDP durations is more extreme in the winter months than in the summer months. Figure 5.11 demonstrates the departure from normality if the model with daily GDP durations is used. A studentized residual is a residual divided by the square root of the mean square error (estimate of variance). If the model is based on average GDP durations instead of daily GDP durations, the assumptions are more nearly satisfied (Figures 5.12 and 5.13). In Figure 5.12, the residuals lie in a horizontal band with one possible outlier (October 96). It appears that the averaging of the daily GDP durations remove enough variability in the data to give constant error variance and normality of residuals.

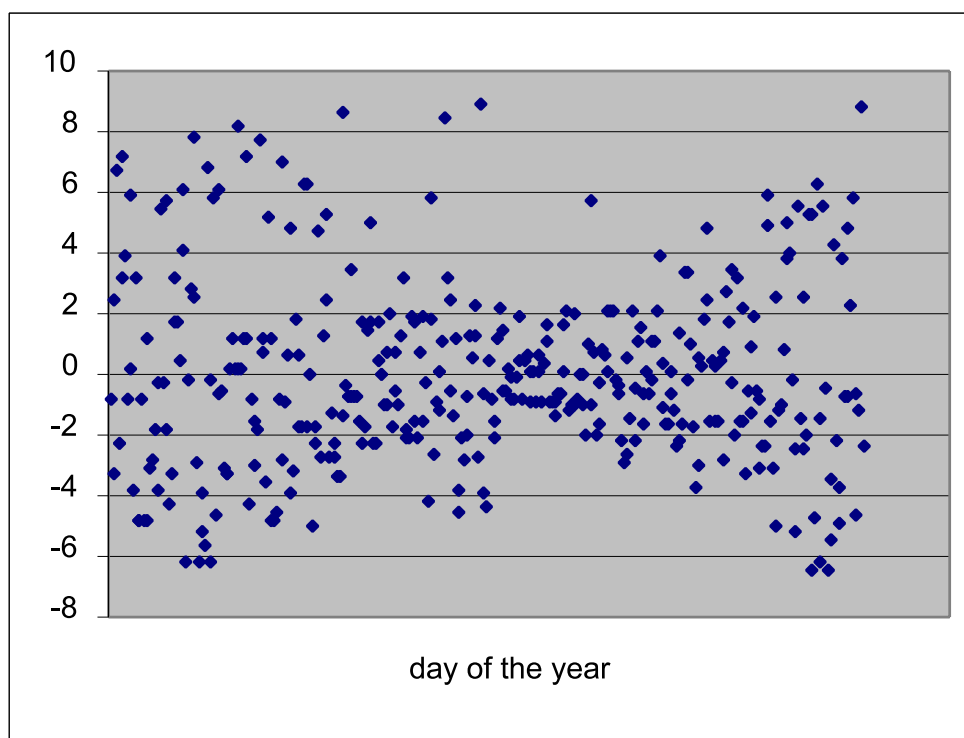


Figure 5.10: Time Plot of Residuals of Daily GDP Durations

An F -test is used to determine if there are statistically significant differences among the means of the months. Using the model involving the average GDP

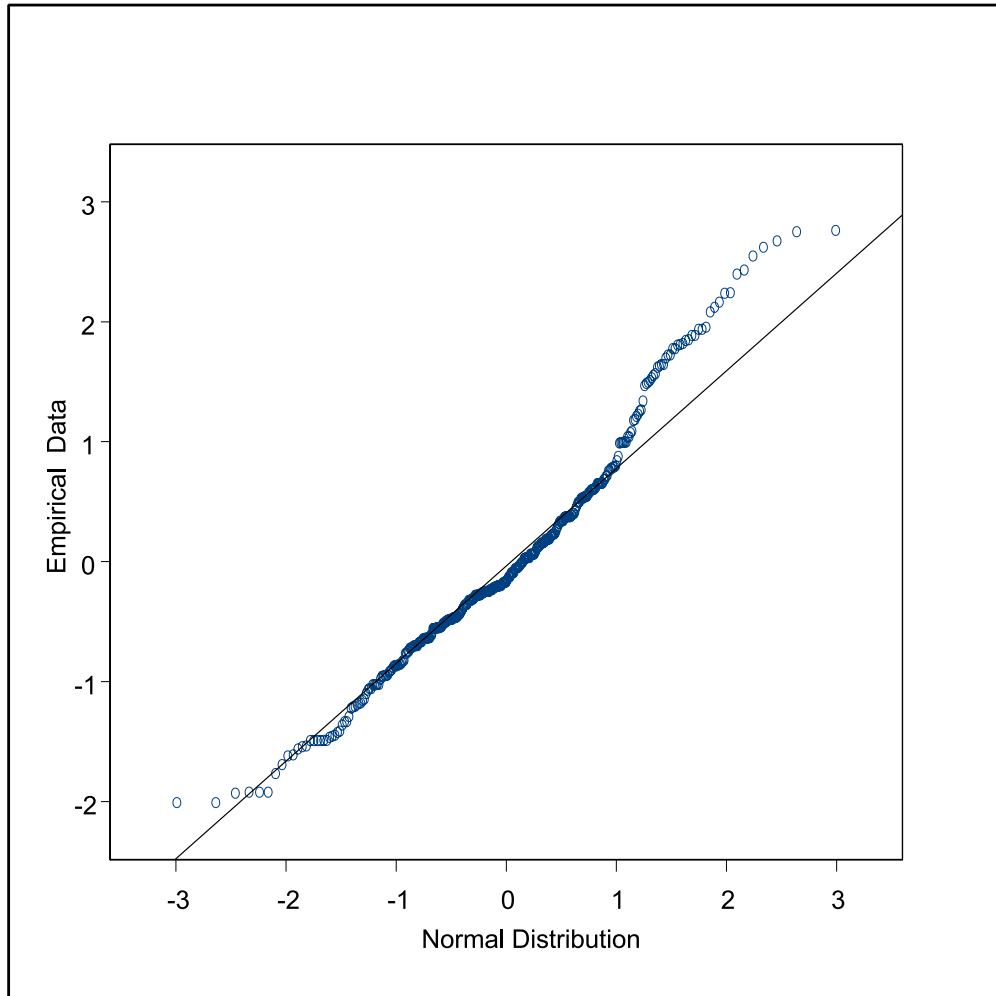


Figure 5.11: QQ Plot of Studentized Residuals of Daily GDP Durations

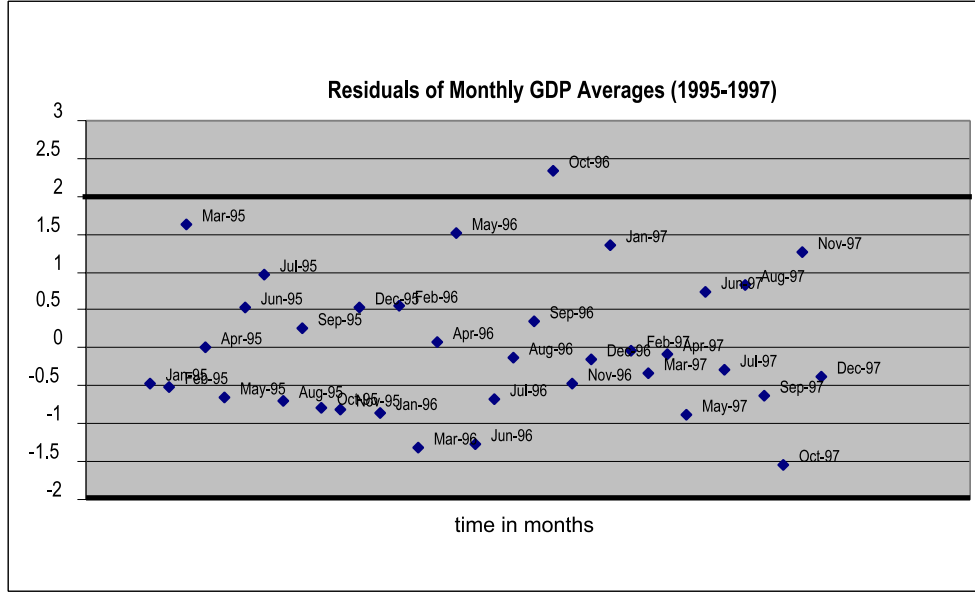


Figure 5.12: Time Plot of Residuals of Monthly GDP Average Durations

durations, the F -test tested the hypothesis that all factor level (monthly) means are equal and resulted in a p -value of .0030, which implies that there does exist some linear function of parameters that is significantly different from 0. In other words, there does exist a significant difference in means of the months.

The set of seasons from the previous sections give some idea of where the differences are. The procedure of multiple comparisons can be used to evaluate a given set of seasons. Multiple comparisons, also known as mean separation tests, tests for equality between two or more factor level means. Each comparison is known as a contrast, L , and is defined as $L = \sum t_j \mu_j$, where t_j satisfies the restriction $\sum t_j = 0$.

Single-Factor ANOVA with multiple comparisons was implemented for the Winter, Spring, Fall and Summer GDP Seasons. Table 5.5 lists the different contrasts (in terms of seasons), their corresponding p -values and standard errors

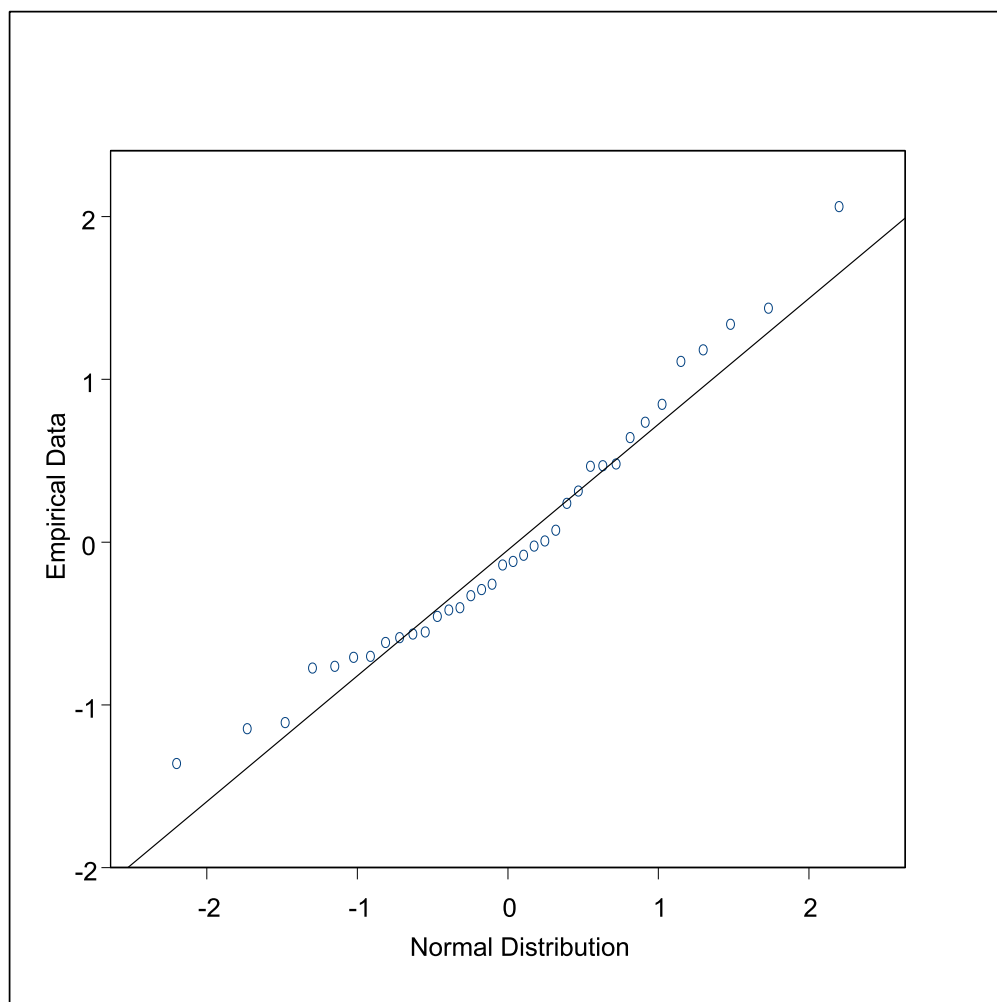


Figure 5.13: QQ Plot of Studentized Residuals of Monthly GDP Average Durations

Contrasts	p-values	standard errors
Jul/Aug vs. Apr-Jun	.0288	1.801
Nov-Mar vs. Apr-Jun	.0015	2.401
Nov-Mar vs. Jul/Aug	.0001	2.750
Nov-Mar vs. Sep/Oct	.0053	2.750
Sep/Oct vs. Jul/Aug	.0388	1.315

Table 5.5: Results of ANOVA Multiple Comparisons Test

that resulted from using the Scheffe' multiple comparisons' procedure. Based on the p -values, there are statistically significant differences between the means of the following pairs of seasons: Jul/Aug and Nov-Mar; Jul/Aug and Sep/Oct; Jul/Aug and Apr-Jun; Nov-Mar and Sep/Oct; and Nov-Mar and Apr-Jun. It should be noted that the hypothesis that there is a difference in the mean of the Sep/Oct season and the mean of the Apr-Jun season could not be rejected. This is due to the fact that they essentially have the same mean, but cannot be placed in the same season because they are separated by the Jul/Aug season, which has a mean that statistically differs from both seasons.

Caution should be taken with this method since the assumptions on the error terms were violated for the actual daily GDP durations. In practice, these assumptions are not often satisfied. When this occurs, a weighted least squares (WLS) method can be used. In ordinary least squares (OLS), the residual sum of squares ($\sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$) is minimized. In WLS, a weighted residual sum of squares ($\sum_{i=1}^n w_i (Y_i - \hat{Y}_i)^2$) is minimized, where w_i is the weight of observation i . Weights are used to adjust the amount of influence of each observation. The normal equations to be solved are

Source	p-value
Model (OLS)	.0030
Model (WLS)	.0171
Month (OLS)	.0015
Month (WLS)	.0136
Year (OLS)	.6629
Year (WLS)	.5900

Table 5.6: F-Test Results for OLS and WLS

$$\beta = (X^T W X)^{-1} X^T W Y$$

where W is a diagonal weight matrix. According to the Gauss-Markov Theorem, estimates from OLS are unbiased and have minimum variance among all unbiased linear estimators. In other words, OLS estimates are best, linear unbiased estimates (BLUE) if the Y_i 's are uncorrelated and homoscedastic. If the Y_i 's are uncorrelated and heteroscedastic, then WLS with weights, $w_i = 1/\text{Var } Y_i$, is BLUE. OLS is performed on GDP data to obtain error variances of the observations to be used as weights in a weighted least squares procedure. WLS is performed using sample derived weights for the purpose of making inferences on the fixed effects of the model. Since we are estimating weights from the empirical data, the resulting WLS estimators may be more efficient than OLS, but are not BLUE. Table 5.6 demonstrates that there is little difference in the p -values for the F -tests using OLS and WLS in the single-factor ANOVA. Using either method, the results state that there do exist statistically significant differences in means related to the month, but not to the year.

To avoid the issue of whether assumptions are satisfied or not for the F -test

to be valid, the mean square ratio (better known as the F -value if normality is satisfied) can be used to evaluate a given set of seasons. The mean square ratio is the ratio of the mean square between groups (seasons) and the mean square within groups (seasons):

$$\frac{\left(\frac{\sum_s n_s (\bar{Y}_{.s} - \bar{Y}_{..})^2}{k - 1} \right)}{\left(\frac{\sum_s \sum_j (Y_{js} - \bar{Y}_{.s})^2}{n - k} \right)}$$

It is desired to have seasons that exhibit homogeneity within seasons and variability between seasons. A mean square ratio that is large confirms that this is the case. The mean square ratios are computed for pairwise contiguous seasons. If the minimum of these values is greater than some large constant, e.g. 10, then the set of seasons is valid. The set of seasons resulting from the set partitioning procedure that minimized the KS statistic using weather data satisfy the mean square ratio criterion. See Table 5.7. For the candidate season set of size 60, the season Mar-Jun will be referred to as the “Rainy” Season, Jul-Sep as the “Summer Weather” Season and, Oct-Feb as the “Heavy Fog” Season. The CPDFs for these seasons will be used as input into the Hoffman-Rifkin model and results will be compared with Command Center plans in Sections 6.1 and 6.3.

Many cost functions were given in Sections 5.3.3 and 5.3.4 to determine sets of seasons and the post analysis in this section is used to determine the best set of seasons. Best refers to a set of seasons where there exists as much intra-season homogeneity as possible and as much inter-season variability as possible. One may be interested in conjecturing a set of seasons instead of performing the complex process (set partitioning IP) of determining a set of seasons. One

	Contiguous Seasons	Mean Square Ratio
n=60	Mar-Jun vs Jul-Sep	14.06
	Jul-Sep vs Oct-Feb	24.39
n=133	May/Jun vs Jul-Sep	11.41
	Jul-Sep vs Oct-Apr	23.33

Table 5.7: Mean Square Ratios of Weather Seasons

example of a set of seasons that could be conjectured is the “standard” set of seasons that are actually referred to as the seasons of the year. The boundaries of these seasons are determined by the dates of the Autumnal Equinox, the Vernal Equinox, the Summer Solstice and the Winter Solstice. December 21st separate the fall and winter seasons; March 21st separate the winter and spring seasons; June 21st separate the spring and summer seasons; and September 21st separate the summer and fall seasons. Since the exact dates of the Winter Solstice, the Vernal Equinox, the Summer Solstice and the Autumnal Equinox change year to year (but always occurring near the end of the months), the seasons are assumed to be January to March, April to June, July to September and October to December. This set of seasons did not satisfy the post analysis criterion because not all mean square ratios of contiguous seasons exceeded 10. See Table 5.8.

The mean square ratios of 1.23 and 1.85 suggests that the within season dissimilarity is basically the same as the between season dissimilarity, which implies that the inter-season variability is small. Since we want (contiguous) seasons to be as different as possible, the “standard” set of seasons is not a very good set.

Contiguous Seasons	Mean Square Ratio
Jan-Mar vs Apr-Jun	1.23
Apr-Jun vs Jul-Sep	13.89
Jul-Sep vs Oct-Dec	12.08
Oct-Dec vs Jan-Mar	1.85

Table 5.8: Mean Square Ratios of “Standard” Seasons

5.3.6 Seasonal “Clustering” Technique

In general, clustering refers to the grouping of objects that are similar. Partitional clustering refers to the decomposition of a data set into a partition or set of disjoint clusters through the minimization of a distance (cost) function. Thus, based on the properties of partitional clustering, a seasonal clustering technique (clustering with an imbedded time series) should have the properties that the clusters chosen are the results of minimizing some cost function, they are disjoint and they contain points that are contiguous.

Usually in a partitional clustering procedure such as K -means, there is a process of searching through the set of all possible clusterings (partitions) to determine the best partition of the data. The set of all possible partitions can be too large so that a local optimization method is needed. In K -Means, a data set of M points are partitioned into K clusters ($P(M, K)$), and the cost criterion is the average squared distance of the observations from their nearest center location. The nearest center location is usually determined by a standard Euclidean distance function. The search procedure is an iterative procedure of assigning data points to clusters that contain their nearest centers, recomputing the center locations and then reassigning the observations until the centers change

by a small (ε) amount.

In a seasonal clustering technique, there is restricted partitioning because the number of possible partitions is reduced due to the restriction that time-order of observations must be maintained. Thus, the set of all possible partitions can be efficiently enumerated, as in a set covering/partitioning procedure. Hence, set partitioning will be used as the search procedure (determination of clusters/seasons) for the seasonal clustering technique. According to discussion in Section 5.3.2, it is known that this IP, whose A matrix possesses the consecutive ones property with wrap around can be solved efficiently. Thus, the seasonal clustering problem can be solved efficiently.

As in partitional clustering, the determination of the best cost (objective) function and appropriate number of clusters are difficult tasks. The cost (objective) function is based only on within season interaction. In the seasonal clustering approach we have proposed, alternative sets of clusters/seasons were generated by the set partitioning model based only on within season interactions. A post-processing step then took into account the between season interactions. See Figure 5.14. Recall from the previous section that the post-processing evaluation criterion is the mean square ratio between contiguous seasons.

The distributions are changing continuously because weather is a continuous process. Because of the continuous process, there are periods between seasons that are transition periods. The evaluation criterion yields the best set of seasons with consideration to these transition periods. It might be desired to create some type of combination of seasons on either side of a transition period to derive the distribution for the transition period.

Seasonal distributions, which have satisfied the mean square ratio criterion,

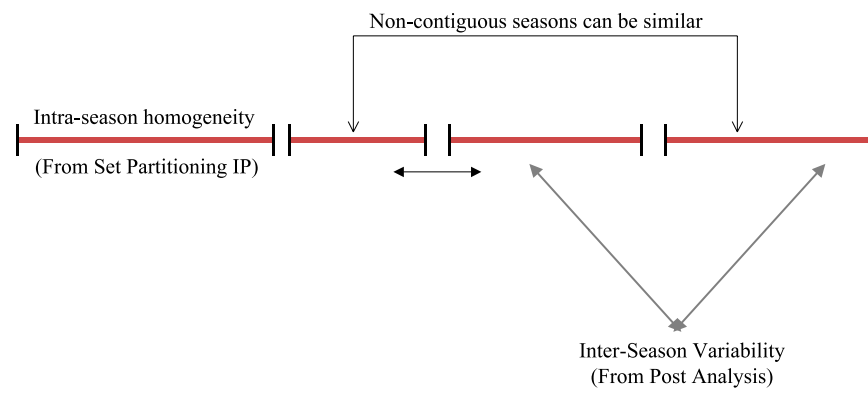


Figure 5.14: Perspective on Seasonal “Clustering”seasonalclustering

may be better operationally than daily distributions because they correspond more to the intuition of the specialists at the ATCSCC. In the case of SFO, the specialists refer to times as “heavy fog” season or “rainy” season. These particular seasons last more than one day or one month. There is also the issue of compactness of seasons. The lengths of resulting seasons that are most appropriate depend on the particular situation being modeled. At SFO, seasons of length one do not make sense because controlling weather conditions such as fog are longer-lived.

5.4 Extension to CPDFs of 2-Parameter ACDs

In Chapter 4, the 2-Parameter ACD is described as a capacity scenario in which both the start time of reduced capacity AND the duration of reduced capacity must both be estimated. Ostensibly, estimating the 2-Parameter ACD is a more challenging problem since both parameters must be estimated. This type of ACD is most appropriate for modeling airports that are plagued with weather conditions that could materialize at any time of the day, e.g. thunderstorms or snow.

Currently, the operational procedures at the ATCSCC cannot accommodate or is not equipped to handle this type of ACD where two parameters need to be estimated. The specialists set a (planned) start time of a GDP to correspond to the start time of reduced capacity. Thus, start time (of reduced capacity) is assumed to be deterministic. In the case of deterministic start times, determining the seasons that vary with start time can be considered a seasonal “clustering” problem. The techniques presented in previous sections could then be used to determine which intervals of start times have the same CPDFs. The problem

now becomes how to find the breakpoints in start times that would determine a change in the CPDFs. CPDFs of 2-Parameter ACDs can be generated in the same way as the CPDFs of 1-Parameter ACDs. In the case of the 2-Parameter ACDs, there would be an overall CPDF for each of the (planned) start times (or intervals of start times) of GDPs. The seasonal “clustering” techniques could again be used to partition the overall CPDF of 2-Parameter ACDs (for each of the GDP start times) into seasonal CPDFs. Hence, we can determine seasonal distributions of durations of IFR conditions given a GDP is planned for every possible start time (or interval of start times). The arrival capacity distribution generator in the ConOps flow chart (Figure 1.1) would need both the season and the start time as inputs to output capacity scenarios (2-Parameter ACDs) with their associated probabilities.

We now consider the case where decision processes and models are available that allow the use of a bivariate distribution whose two parameters are start time and duration of reduced capacity. In this case, instead of calibrating a CPDF of 1-Parameter ACDs, we must calibrate a CPDF of 2-Parameter ACDs. In principle, the same general approach can be applied, but the number of bins will increase substantially. Thus, to get more accurate estimates, more empirical data would be required. In addition, the seasonal clustering technique would become more complex. If the data supports the hypothesis that start time of reduced capacity is independent of duration of reduced capacity, then the problem could be solved by estimating two CPDFs of 1-Parameter ACDs. Each of these could be estimated using the methods of this thesis.

Chapter 6

Decision Models and Evaluation of ACD Models

6.1 The Hoffman-Rifkin Static Stochastic Ground Holding Model

In their theses, Ryan Rifkin (MIT)[35] and Robert Hoffman (University of Maryland)[21] developed integer programming models to address the static stochastic version of the ground holding problem (GHP). (For a succinct description of the model, see [5]). Recall that the GHP is the problem of determining an optimal balance between the amount of delay to assign to flights to be taken on the ground during a GDP and the amount of expected airborne delay. The Hoffman-Rifkin static stochastic ground holding problem (H-R) [5] is formulated as:

$$\text{Minimize} \quad \sum_{t=1}^T c_g G_t + \sum_{q=1}^Q \sum_{t=1}^T c_a p_q W_{q,t}$$

subject to

$$A_t - G_{t-1} + G_t = D_t \quad t = 1, \dots, T+1 \quad (1)$$

$$G_0 = G_{T+1} = 0$$

$$-W_{q,t-1} + W_{q,t} - A_t \geq -M_{q,t} \quad t = 1, \dots, T + 1, \quad (2)$$

$$q = 1, \dots, Q$$

$$W_{q,0} = W_{q,T+1} = 0$$

$$A_t \in Z_+, W_{q,t} \in Z_+, G_t \in Z_+ \quad (3)$$

The objective of the H-R model is to minimize the sum of the costs of assigned ground delay and the costs of expected (unplanned) airborne delay. The decision variables, A_t , represent the number of flights that should arrive at the airport in time period t with no airborne delays. One can think of A_t as a planned AAR (PAAR) during time period t . Also, $W_{q,t}$ is the number of flights delayed in the air from time period t until a subsequent time period under scenario q and $M_{q,t}$ is the arrival capacity during time period t under scenario q . A sequence of $M_{q,t}$ for the whole time horizon, T , is one possible capacity scenario q or ACD. Recall the various forms of ACDs as discussed in Chapter 4. The interpretation of G_t has certain subtleties. It can be thought of as the number of flights delayed on the ground from time period t to $t + 1$. However, here time is measured relative to the time at which these flights would arrive at the airport. The actual time at which the delay would be taken is determined by choosing the specific flights to be delayed and then subtracting the appropriate (flight-specific) en-route times. Constraint set (1) states that all flights that are predicted to arrive in time period t (demand or D_t) or were delayed on the ground from the previous time period (G_{t-1}) must arrive in the current time period (A_t) or be delayed on the ground until a subsequent time period (G_t). Constraint set (2) states that under scenario q , all flights scheduled to arrive in the current time period or that are air delayed

from a previous time period ($W_{q,t-1}$) must be air delayed until a subsequent time period or must arrive in the current time period ($M_{q,t}$).

The inputs into the H-R model are: the number of predicted arrivals or demand for each time period t (D_t), the cost of ground delaying one flight for one time period (c_g), the cost of one period of airborne delay of a single flight (c_a), and Q capacity scenarios (ACDs) with associated probabilities, p_q . The output of the model is the number of flights that should land in a given time period t , A_t , i.e., the number of arrival slots that should be made available in each time period t .

The H-R model assumes that all ground delay assigned under a particular output scenario is realized, deterministic and independent of the scenario. The model does not take into consideration the dynamic changes that may occur if one scenario is planned and another occurs. Thus, the model does not give the flexibility to make changes in assigned ground delay as forecasted weather conditions change.

6.2 Adjusting Assigned Ground Delay for Dynamically Changing GDPs

The H-R model attempts to capture the stochastic nature of weather through the probabilistic distribution of capacity scenarios. It outputs a particular scenario based on demand, the air to ground cost ratio and the probability of the scenario. It does not capture the existing ability to dynamically change assigned ground delay as (predicted) conditions change. For example, if a dissipation of predicted weather occurs, then it may be possible to reduce previously assigned ground

delay. Thus, some assigned ground delay may be recovered if a GDP is canceled due to dissipation of inclement weather. Alternatively, if the duration of poor weather is longer than expected, then the GDP can be revised/extended, hereby assigning additional ground delays.

6.2.1 Shortened Reduced Capacity Due to Canceled GDP

Here we consider the case where a GDP is canceled, thus the duration of reduced capacity is shortened. Depending on how far a flight's (controlled) departure time is from the cancellation time of a GDP, the flight can recover some or all of its assigned ground delay. Suppose a flight f had an original estimated time of departure (OETD) of 12:30, but under a GDP, it was given a controlled time of departure (CTD) of 1:15. Now suppose that the inclement weather clears such that the GDP is canceled at 12:15. Since flight f is still on the ground at this cancellation time and full capacity has been restored, it is allowed to take off as soon as possible. However, a number of factors, e.g., the status of the passengers, might delay the time at which the flight is able to depart. For example, its actual runway time of departure (ARTD) might be 12:45. In this case, some of its assigned delay (CTD-OETD) is recovered. The actual ground delay (GD) realized is ARTD-OETD. In this scenario:

- Assigned GD = CTD - OETD = 1:15 - 12:30 = 45 minutes
- Actual GD = ARTD - OETD = 12:45 - 12:30 = 15 minutes
- GD Recovered = Assigned GD - Actual GD = CTD - ARTD = 1:15 - 12:45 = 30 minutes

Flights whose CTDs are prior to the GDP cancellation time (CNXTime) will incur all of their assigned GD. Thus, only flights that are controlled to depart after the CNXTime can recover some of their assigned GD. If a flight's OETD is before the CNXTime, then the amount of assigned GD that is available for recovery is CTD-CNXTime. See Figure 6.1.

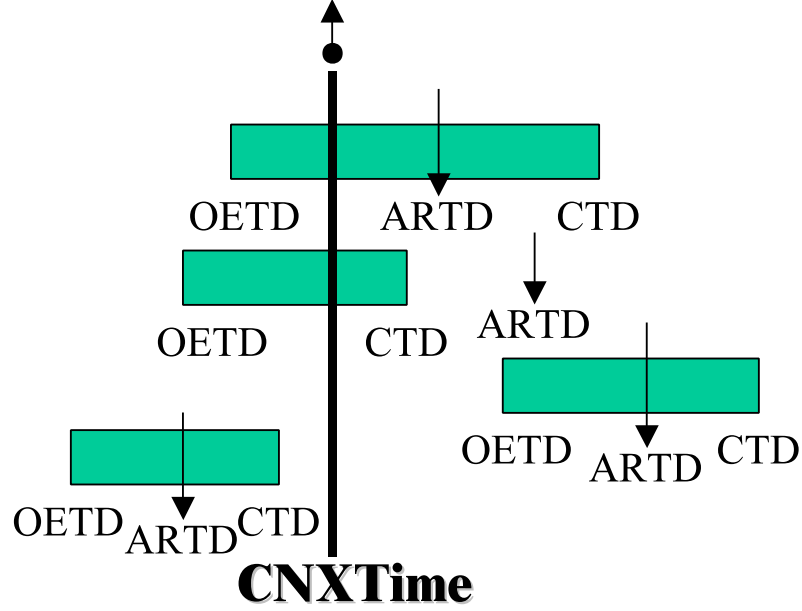


Figure 6.1: Ground Delay Available for Recovery (Recoverable GD)

$$\frac{ARTD - \max(OETD, CNXTime)}{\min(CTD - OETD, CTD - CNXTime)}$$

The above percentage can be greater than 1 if a flight's ARTD is greater than its CTD. This means that the flight incurred extra delay (possibly) unrelated to the GDP. In our data analysis, such flights are assumed to have incurred 100% of their assigned GD. To determine the percentage of recoverable GD that was recovered, we subtract the above value from 1:

$$1 - \left[\frac{ARTD - \max(OETD, CNXTime)}{\min(CTD - OETD, CTD - CNXTime)} \right]$$

Using the information in the ADL files (Section 3.4) for all flights scheduled to arrive at SFO on all days in 1998 that a GDP was planned and run during the morning hours, the percentage of recoverable GD recovered as a function of a flight's CTD minus CNXTime is calculated. Figure 6.2 graphs the change in this percentage as a function of CTD-CNXTIME. Our prior conception was that the functional relationship should be logarithmic. (Other functions were fitted to the data, but the log yielded the largest R^2 value). However, it is difficult to fit a good function (logarithmic) to the data. Since the log function can grow above the value of 100, a more appropriate function may be one that has an asymptote at $y = 100$.

The graph in Figure 6.2 contains data points whose values (percentages) were fixed at 0 if they were originally negative and fixed at 100 if they originally exceeded 100. As stated previously, any flight whose percentage is negative is indicative of extra delay being incurred possibly from mechanical delays or other types of problems. Flights whose percentages exceed 100 have made up more delay than the oversimple model indicated was possible. Those flights might have departed prior to their CTDs or could have been exempted from the GDP.

Several methods for filtering the data were implemented. Using the original percentages, the data was filtered by removing those data points whose percentages were below -10% or above 110% (Figure 6.3). The data was also filtered by considering only those flights whose assigned GD was no more than 2 hours (Figure 6.4). Another method of filtering is to calculate percentage of recovered GD for those flights whose CTDs are no more than 3 hours after the CNXTime (Figure 6.5).

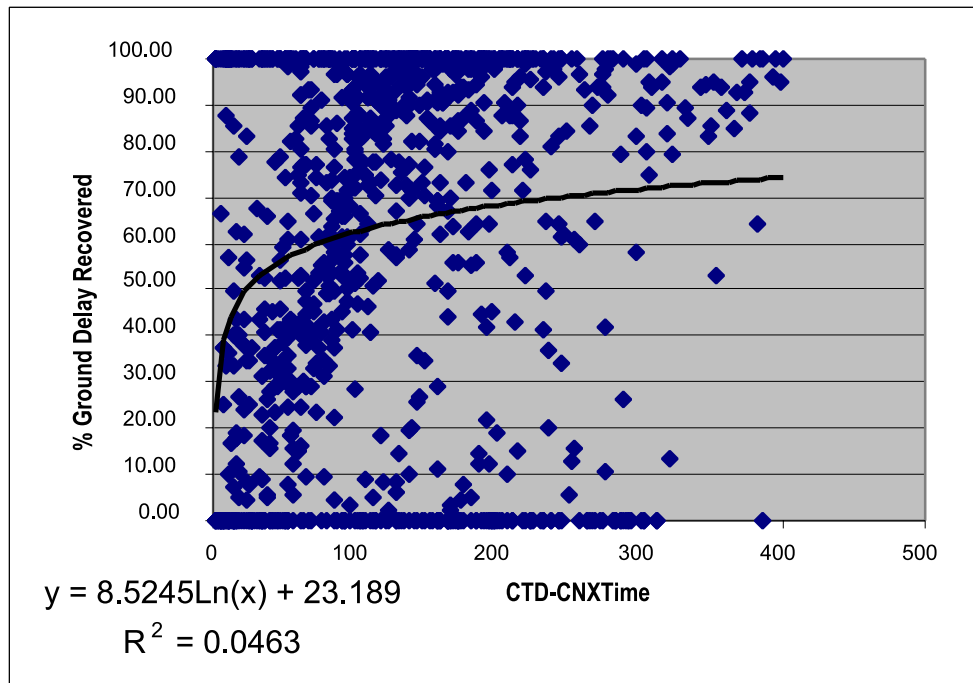


Figure 6.2: Plot of Percentage of Ground Delay Recovered as a Function of Difference of (Controlled) Departure Times from Cancellation Times of GDPs

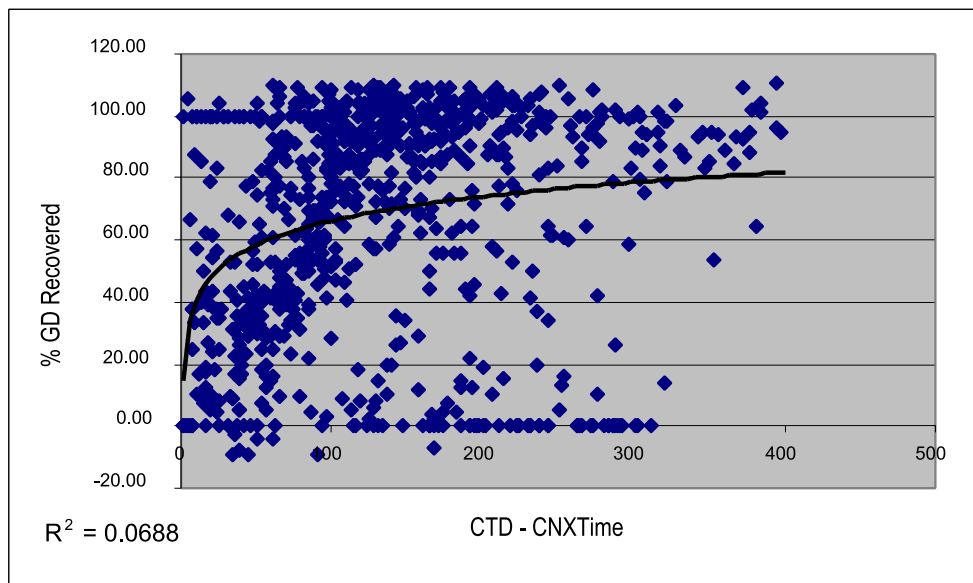


Figure 6.3: Filtered Percentage of GD Recovered for Percentages [-10,110]

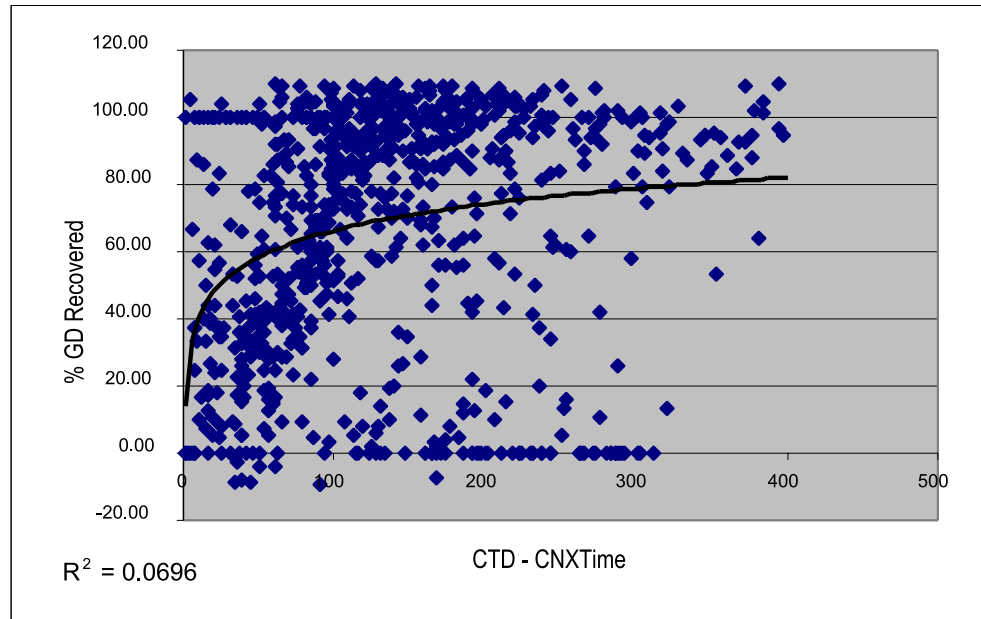


Figure 6.4: Percentage of GD Recovered for Assigned GD < 2 Hours

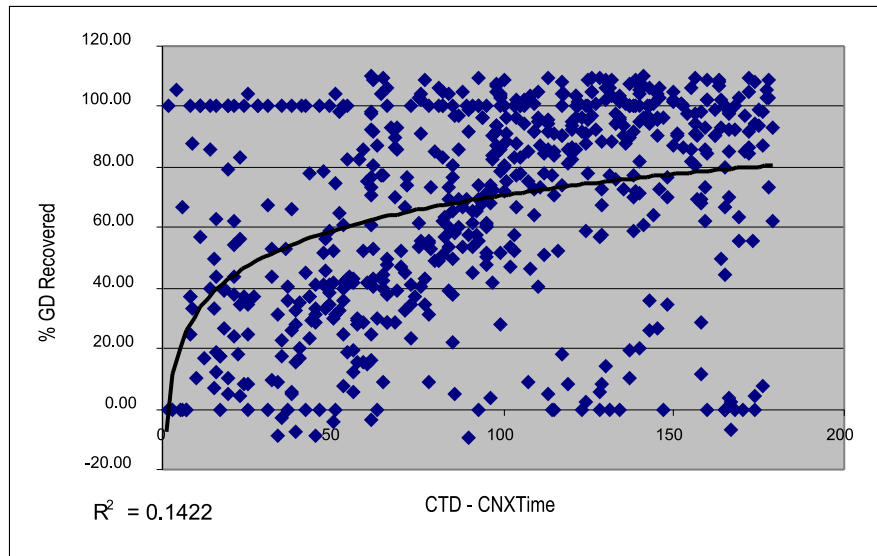


Figure 6.5: Percentage of GD Recovered for Flights Whose (Controlled) Departure Times Are Less Than 3 Hours After Cancellation Times of GDPs

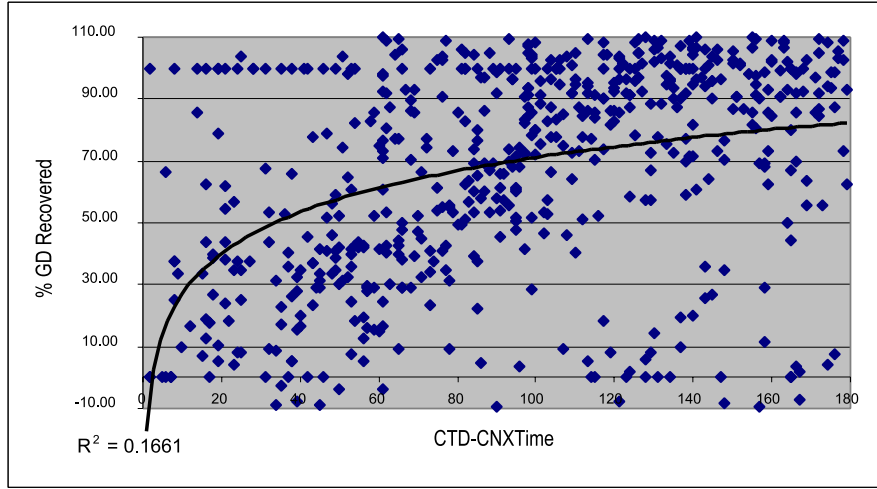


Figure 6.6: Plot of Percentage GD Recovered After All Filtering

Filter	Reason		
CTD - OETD = 0	Makes denominator 0	1143	1082
CTD - OETD \neq 0	No control before scheduled	1082	1040
CTD - OETD $\in [11,120]$	GD no more than 2 hours	1040	971
CTD - CNXTIME < 180	Depart less than 3 hrs after CNX	971	761
% GD Recovered $\in [-10,110]$	(see section)	761	587

Table 6.1: Filtering Criteria

No one filtering method resulted in a set of data to which a good logarithmic function could be fitted (low R^2), so a combination of filtering methods was used. Table 6.1 lists the various filtering criteria, the reasons the criteria are reasonable, and the number of observations before and after the particular filtering is performed. Figure 6.6 is the result of the combination of filtering techniques. Though the R^2 value (amount of total variation explained by the function) increases, it is still not a very good fit. Thus, averages of percentages of GD recovered are

Time Intervals	% GD Recovered
0-30 mins	42 %
31-60 mins	42 %
61-90 mins	66 %
91-120 mins	78 %
≥ 120 mins	80 %

Table 6.2: Actual Percentages of GD Recovered in a Canceled GDP

calculated in 30-minute intervals (Table 6.2). It is assumed that any flight whose CTD is no more than 30 minutes after the CNXTime recovers none (0%) of its assigned GD and any flight whose CTD is more than 2 hours after the CNXTime recovers all (100%) of its assigned GD. Table 6.3 gives the average percentages for each 30-minute bucket for this case. These assumptions will be observed and the respective percentages will be used to modify the output of the H-R model. The amount of assigned (recoverable) GD in a canceled GDP is adjusted by the (recoverable) amount that is recovered,

$$\text{Recoverable GD realized} = \text{Assigned Recoverable GD} - (\% \text{ GD Recovered}) * (\text{Assigned Recoverable GD}).$$

6.2.2 Lengthened Reduced Capacity Due to Revised GDP

With the emergence of CDM came the flexibility to “revise” different parameters of the GDP as conditions (weather or demand) change. If a GDP is revised/extended due to the worsening of weather conditions, then the originally assigned ground delay is modified and, thus depends on the scenario. The H-R model assumes that ground delay is deterministic and that, if reduced capacity

Time Intervals	% GD Recovered
0-30 mins	0 %
31-60 mins	40.80 %
61-90 mins	65.20 %
91-120 mins	77.15 %
≥ 120 mins	100 %

Table 6.3: Average Percentages of GD Recovered in a Canceled GDP

lasts longer than the duration in the planned scenario, then all “extra delay” is in the form of airborne delay. Thus, it overestimates airborne delay.

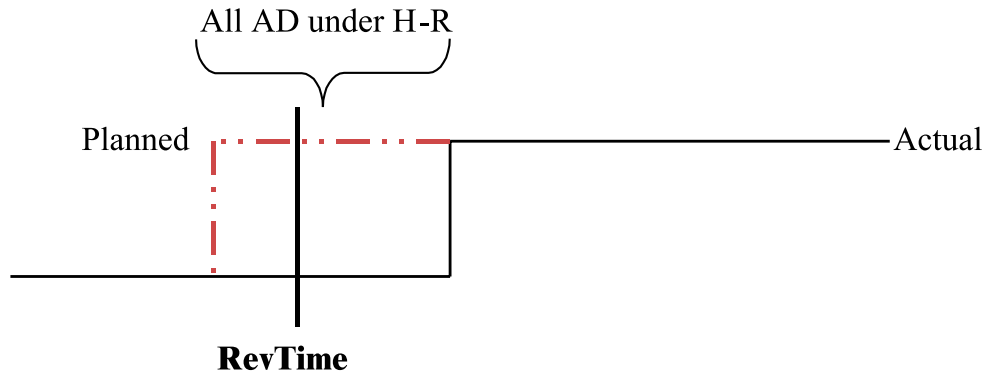


Figure 6.7: Additional GD under Revised GDP

If a flight’s CTD is before the revised time (RevTime) of the GDP, then it may indeed incur unplanned airborne delay (AD), but if the flight’s CTD is after the RevTime, then it should incur only extra GD.

In a GDP, flight delays are initially calculated by setting a controlled time of arrival (CTA). Assigned GD is set equal to CTA-OETA, and GD is then added to the OETD to determine the CTD. Flights are exempted from the revised portion

of the GDP if

$$\text{CTD} - \text{RevTime} < 0$$

and the airborne delay they may incur is calculated by subtracting their CTAs under the planned scenario from their CTAs under the actual scenario. On the other hand, flights whose CTDs satisfy

$$\text{CTD} - \text{RevTime} > 0$$

can be assigned additional delay on the ground. The additional delay is calculated just as stated above for the other case, but all the delay is taken on the ground. Hence, assigned GD from the H-R model needs to be adjusted appropriately for the flights in a revised GDP.

6.3 Analysis and Computational Results

6.3.1 H-R Results (M_PAAR) for Seasonal CPDFs

Recall that the required inputs into the H-R model are the demand or predicted arrivals for each time period over a given (discretized) time horizon, the capacity scenarios (1-Parameter ACDs) with their associated probabilities, the cost of one unit of (assigned) ground delay, and the cost of one unit of expected airborne delay.

Actual GDPs during 1998 were run through the H-R model to determine the PAAR from the model (M_PAAR) over a given time period. ADL files (Section 3.4) were used to determine the aggregate demand for each time period 4 hours in advance of the planned start time of the GDP. Specialists at the ATCSCC

plan a GDP at least 4 hours in advance based on forecasted weather conditions, predicted demand and capacity. In the ADLs, there are only 7 periods (hours) of predicted demand 4 hours in advance of the start time of the GDP. Thus, each ACD contain only 7 periods of capacity. There are a total of 8 input capacity scenarios (Table 6.4). The associated probabilities used depend on the seasonal CPDFs of choice. For analysis purposes, the seasonal CPDFs used are those resulting from the set partitioning method that minimized differences in EDFs implemented on weather data (Section 5.3.3). These seasons were referred to as the “Heavy Fog” (Oct-Feb) season, the “Rainy” (Mar-Jun) season and the “Summer Weather” (Jul-Sep) season. These seasonal CPDFs were chosen for analysis because the F -values between contiguous seasons satisfy the mean square ratio criterion for a good set of seasons (Section 5.3.5). The associated probabilities for the 1-Parameter ACDs in the seasonal CPDFs (Figures 6.8, 6.9, 6.10) are conditioned appropriately for the inclusion of the 0-hour reduced capacity ACD from the GDP data. The costs of one unit (minute) of ground delay and air delay were based on a study by the Air Transport Association (March 2, 2000) and reported by Metron, Inc. [1] The study concluded that the cost of one minute of delay at the gate is \$24.30, the cost of one minute of taxi-out delay is \$30.47 and the cost of one minute of airborne delay is \$47.64. Based on these values, one unit of airborne delay costs 1.96 times more than one unit of ground delay. The H-R model was run with $c_g = 1$ for three alternative airborne delay factors: $c_a = 1.5$, $c_a = 2.0$, and $c_a = 2.5$. The M_PAAR results are given in Table 6.4. It is observed that the output scenario corresponds to one of the input scenarios for any given season. Each day of a given season has the same resulting scenario and this may be due to consistent day-to-day projected demand levels. Based on the

# of Hours of Reduced Capacity	1-Parameter ACD
0	45 45 45 45 45 45 45
1	30 45 45 45 45 45 45
2	30 30 45 45 45 45 45
3	30 30 30 45 45 45 45
4	30 30 30 30 45 45 45
5	30 30 30 30 30 45 45
6	30 30 30 30 30 30 45
7	30 30 30 30 30 30 30

Table 6.4: Capacity Scenarios (Inputs into the H-R Model)

Weather Season	$c_a = 1.5$	$c_a = 2.0$	$c_a = 2.5$
Oct-Feb	3 hours	4 hours	5 hours
Mar-Jun	2 hours	3 hours	4 hours
Jul-Sep	2 hours	2 hours	3 hours

Table 6.5: Results of the H-R Model (Number of Hours of Reduced Capacity)

assumption that the output scenario will be one of the input scenarios, a general decision model used to cost out scenarios for a given plan is presented in Section 6.4.

6.3.2 Procedure for Comparing Planned and Actual Capacity Scenarios

Let F be a set of flights (f) scheduled to arrive at a congested airport. We denote by M_PAAR the PAAR based on results from the H-R model, by CC_PAAR the

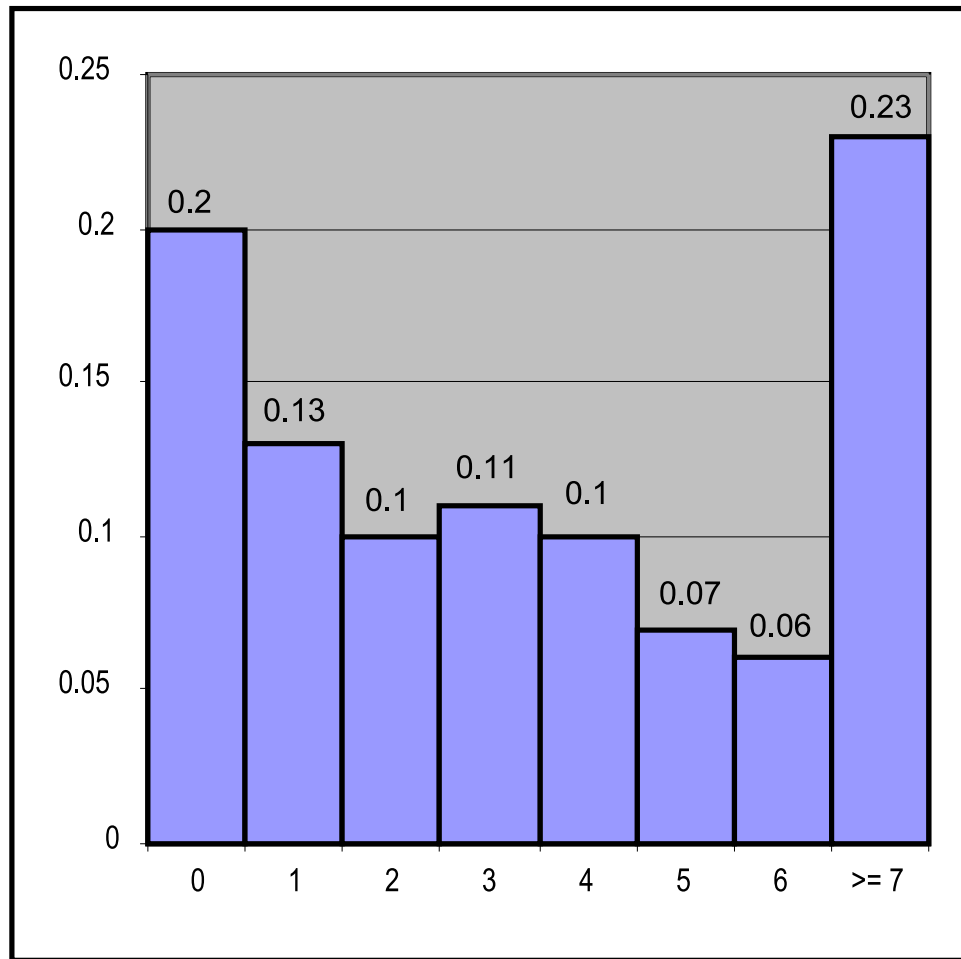


Figure 6.8: Conditional Frequency Histogram for “Heavy Fog” Season

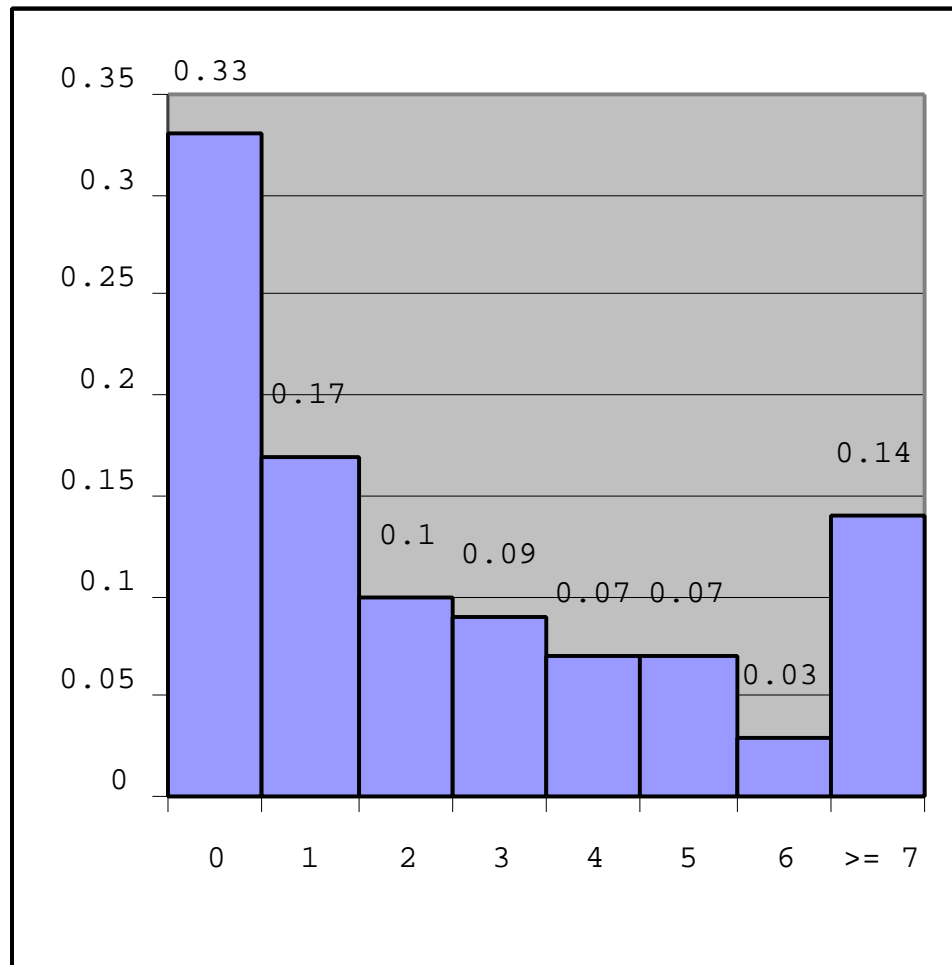


Figure 6.9: Conditional Frequency Histogram for "Rainy" Season

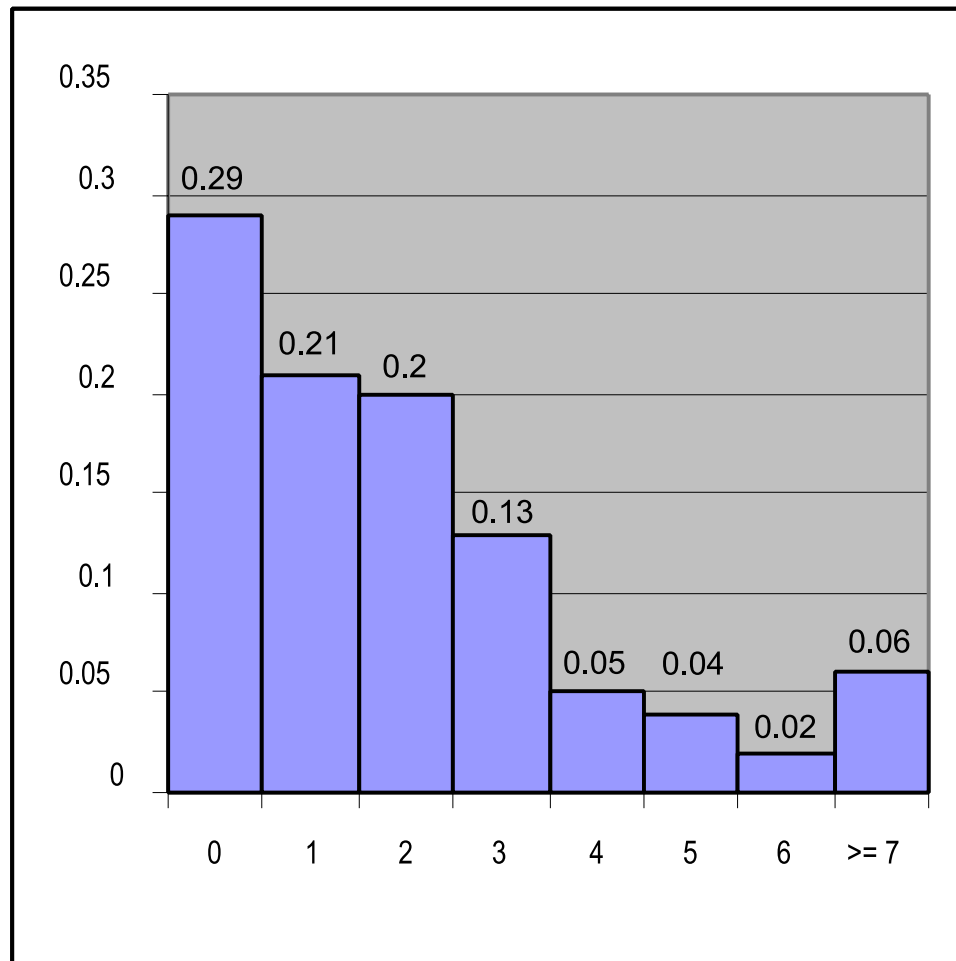


Figure 6.10: Conditional Frequency Histogram for "Summer Weather" Season

PAAR based on the Command Center's planned duration of the GDP, and by ACT_PAAR the PAAR based on the actual duration of the GDP (baseline).

Given the planned and actual durations of reduced capacity, the procedure assigns CTAs or virtual arrival slots to the flights that fall into these durations or thereafter. The procedure proceeds by calculating the amount of assigned ground delay under each "plan", adjusting the amounts appropriately depending on whether the plan is greater than the actual or vice versa, and finally calculating the total weighted delay of a given plan in comparison to the total weighted delay of an actual "plan".

For each $f \in F$, the data fields to be used in the model are:

M_CTA^f is the controlled time of arrival under M_PAAR.

CC_CTA^f is the controlled time of arrival under CC_PAAR.

ACT_CTA^f is the controlled time of arrival under ACT_PAAR.

$OETA^f$ is the original estimated time of arrival.

CTD^f is the actual controlled time of departure.

GD_{xxx}^f is the amount of assigned GD under plan xxx .

AD_{xxx}^f is the amount of AD under plan xxx .

TGD_{xxx} is the total amount of assigned GD for all flights under plan xxx .

TAD_{xxx} is the total amount of (weighted) AD for all flights under plan xxx .

TWD_{xxx} is the total weighted delay (sum of GD and weighted AD) for all flights under plan xxx .

The time parameters are:

t_0 is the start time of the GDP (GDP_Start) under any of M_PAAR, CC_PAAR, ACT_PAAR.

T is the entire time horizon for time period t , $t = t_0, \dots, T$.

t_1^M is the planned end time of the GDP under M_PAAR.

t_1^{CC} is the planned end time of the GDP under CC_PAAR.

t_1^{ACT} is the end time of the GDP under ACT_PAAR, i.e. the actual end time of reduced capacity.

$t_0 \leq t \leq t_1^{xxx}$ is the time interval of reduced capacity under plan xxx .

The overall approach to comparing the total weighted delay incurred under M_PAAR and under CC_PAAR to total weighted delay under ACT_PAAR is as follows:

Main Routine

Step 1:

Determine M_PAAR_t for $t_0 \leq t \leq t_1^M$;

Determine CC_PAAR_t for $t_0 \leq t \leq t_1^{CC}$;

Determine ACT_PAAR_t for $t_0 \leq t \leq t_1^{ACT}$.

Step 2:

(1) Determine set of flights to include in GDP:

- Under M_PAAR: $\{f \in F : t_0 \leq OETA^f \leq t_1^M\}$
- Under CC_PAAR: $\{f \in F : t_0 \leq OETA^f \leq t_1^{CC}\}$
- Under ACT_PAAR: $\{f \in F : t_0 \leq OETA^f \leq t_1^{ACT}\}$

(2) Initialize M_CTA^f , CC_CTA^f , and ACT_CTA^f to $OETA^f$ for $OETA^f = \text{GDP_START}(t_0)$, for all flights that satisfy $OETA^f \geq t_0$.

Step 3:

Assign flights CTAs or virtual slots under each of M_PAAR, CC_PAAR, ACT_PAAR (See Subroutine).

Step 4:

Determine amounts of (initial) assigned GD for each flight f under each of M_PAAR, CC_PAAR, and ACT_PAAR.

- $GD_{M_PAAR}^f = M_CTA^f - OETA^f$;
- $GD_{CC_PAAR}^f = CC_CTA^f - OETA^f$;
- $GD_{ACT_PAAR}^f = ACT_CTA^f - OETA^f$.

Step 5:

Adjust assigned GD and determine AD when planned duration is shorter or longer than actual duration (as discussed in Sections 6.2.1 and 6.2.2).

If $M_PAAR < ACT_PAAR$ (planned too short), then GDP may be revised and

$$\begin{aligned} \text{for } CTD^f > RevTime, \quad GD_{M_PAAR}^f &= GD_{M_PAAR}^f + (ACT_CTA^f - M_CTA^f); \\ \text{for } CTD^f < RevTime, \quad AD_{M_PAAR}^f &= ACT_CTA^f - M_CTA^f. \end{aligned}$$

If $M_PAAR > ACT_PAAR$ (planned too long), then GDP may be canceled, and the adjusted GD is:

$$GD_{M_PAAR}^f = GD_{M_PAAR}^f - GD_{Recoverable}^f * \%GD^f_{Recovered}(CTD^f - CNXTime).$$

(Similarly, assigned GD can be adjusted and AD calculated under CC_PAAR.)

Step 6:

Determine Total Assigned GD under M_PAAR (TGD_M) by

$$TGD_M = \sum_{f \in F} GD_{M_PAAR}^f.$$

Determine Total Airborne (weighted) Delay (TAD_M) under M_PAAR for Revised/Extended GDP by

$$TAD_M = (\sum_{f \in F} AD_{M_PAAR}^f) * c_a.$$

(Similarly, total assigned GD and weighted AD can be calculated under CC_PAAR.)

Step 7:

Calculate Total Weighted Delay Under M_PAAR (TWD_M) by summing total GD and total weighted airborne delay,

$$TWD_M = TGD_M + TAD_M.$$

(Similarly, total weighted delay can be computed for CC_PAAR.)

Step 8:

Calculate Average Total Weighted Delay over representative sample of GDP Days under M_PAAR:

$$AvgTWD_M = \frac{1}{n} \sum_{d=1}^n Sum_d(TGD_M, TAD_M).$$

(Similarly, average total weighted delay can be computed for CC_PAAR.)

Step 9:

Compare average total weighted delay under M_PAAR and CC_PAAR to ACT_PAAR according to:

AvgTWD_M - Avg(TGD_ACT) compared to

AvgTWD_CC - Avg(TGD_ACT).

Subroutine:

If $M_PAAR_t = 30$ (flights per time period t), then the set of virtual slots, S, is $\{\max(OETA^f, t_0) = M_CTA_0, \max(OETA^f, M_CTA_0 + 2), \max(OETA^f, M_CTA_0 + 4), \dots, t_1^M\}$.

If $M_PAAR_t = 45$ (for $t \geq t_1^M$), then the set of virtual slots, S, is $\{\max(OETA^f, M_CTA_{t_1^M} + 1), \max(OETA^f, M_CTA_{t_1^M} + 3), \max(OETA^f, M_CTA_{t_1^M} + 4), \max(OETA^f, M_CTA_{t_1^M} + 6), \dots, T\}$.

(Similarly, flights can be assigned CTAs under CC_PAAR and ACT_PAAR.)

$AvgTGD_M = 7284$	$AvgTGD_CC = 8914$	$Avg(TGD_ACT) = 6875$
$AvgTAD_M = 2417$	$AvgTAD_CC = 1314$	$Avg(TAD_ACT) = 0$
$AvgTWD_M = 9007$	$AvgTWD_CC = 9850$	$Avg(TWD_ACT) = 6875$

Table 6.6: Total Weighted Delay of H-R Plans Vs Command Center Plans

6.3.3 Comparisons of H-R Results with Command Center Plans

GDP days used in the analysis of M_PAAR and CC_PAAR are days in 1998 whose ADL files did not contain any unreliable data due to temporary lapses in the data stream over the CDMnet and whose initial PAAR is 30 flights per hour. Due to these restrictions, there were not many GDP days available for analysis. Therefore, a representative sample, a sample that represented the overall outcomes of the GDPs, was chosen. This representative sample is based on the breakdown of the types (outcomes) of GDPs in 1998 at SFO. (See Figure 6.11) The representative sample includes 11 GDPs, of which 6 are canceled, 4 are revised/extended and 1 is run out. Run out means that the original planned duration of the GDP is the same as the actual duration of the GDP.

Average total weighted delay was determined for M_PAAR, CC_PAAR, and ACT_PAAR according to the procedure in Section 6.3.2. Table 6.6 gives a summary of these results. See Appendix D for the detailed breakdown of results.

According to Table 6.6, both the average of total GROUND delay and average of total weighted delay under M_PAAR are less than these values under the Command Center's plan. The H-R model produces less total weighted delay than the Command Center's plan and this is confirmed because the difference between the average total weighted delay under M_PAAR and the average total

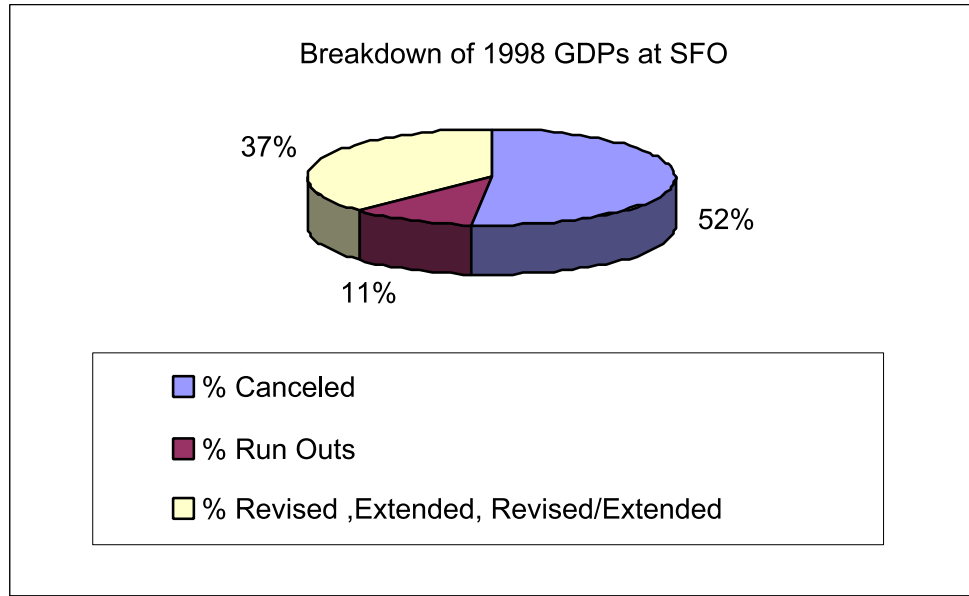


Figure 6.11: Percentage of Outcomes of 1998 Morning GDPs

(ground) delay under ACT_PAAR is less than the difference between the average total weighted delay under CC_PAAR and the average total (ground) delay under ACT_PAAR:

$$AvgTWD_M - Avg(TWD_{ACT}) = 9007 - 6875 = 2132 \text{ minutes, and}$$

$$AvgTWD_{CC} - Avg(TWD_{ACT}) = 9850 - 6875 = 2975 \text{ minutes.}$$

Overall for the representative sample of GDP days in 1998, there could have been a savings of 843 delay minutes if our model had been used. Since there is a reduction in delays, our model is capturing what it should. In general, our model results in shorter planned durations of GDP than is currently employed. It appears that the best recommendation to the ATCSCC is to plan shorter programs since there is a reduction in delay minutes. Though delay is reduced, planning shorter programs may not be operationally feasible. See Section 6.5 for

a more detailed discussion.

6.4 Proposed General Decision Model

In Section 6.1, the resulting (output) ACD was one of the input capacity scenarios. Thus, the number of possible solutions is very small, so that a strategy of enumerating all possible solutions and costing out each one is computationally feasible. We propose this approach as a general decision model and show how to cost out each solution. This decision model is an improvement over the H-R model because the amount of ground delay assigned is adjusted depending on if the plan is too short or too long. In addition, airborne delays are more accurately estimated.

Section 6.3.2 gave an algorithm for comparing an M_PAAR to a CC_PAAR. This approach can be generalized to compare a planned duration of reduced capacity (PAAR) to all possible actual durations of reduced capacities (AAR). Thus, this decision model is calculating a cost for a particular decision (solution) based on an outcome of one of the scenarios. The algorithm for this costing out process is:

For each proposed solution i ,

For each scenario j ,

For each possible cancellation/revised time t_k ,

Calculate cost_{ijk} , the cost of using solution i when scenario j occurs

$$([\sum_{f \in F} (GD_s^f + AD_s^f)]);$$

$$\text{cost}_i = \text{cost}_i + p_j * q_k * \text{cost}_{ijk};$$

Output $i^* = \text{argmin} \{\text{cost}_i\}$.

6.5 Benefits Analysis and Future Vision of CDM

With the inception of initial CDM prototype operations in January 1998 came new procedures and methodologies to make decision-making more distributive between the ATCSCC and the airlines and to give everyone a common view of the airspace. CDM with its new procedures has made a significant impact on the effectiveness of GDPs and thus has decreased the amount of unnecessary delay. The work in this thesis is principally aimed at new decision procedures. However, as a by-product, we were able to estimate some impacts of CDM procedures on system performance.

6.5.1 Increase in GDP Effectiveness under CDM

Analyses were performed on GDPs prior to CDM and after the inception of CDM to gauge the benefits and impact of CDM on GDPs. CDM went into initial prototype operations in January 1998 at SFO and EWR. It was hypothesized originally that any cancellation of a GDP is an indication of an inappropriately planned program, and therefore, the number or percentage of canceled programs should decrease as time passed from Pre-CDM to CDM phases.

Using PROC CATMOD in SAS, loglinear analysis was performed on cell frequencies in contingency tables where there are no a priori distinctions between response and predictor variables. All variables are treated as response variables. Based on analysis and comparison of counts of canceled GDPs between 1997 and 1998 (for ATL, BOS, EWR, ORD, SFO and STL), there exists a statistically significant interaction between airport/ # of GDPs canceled and airport/ year/ # of GDPs canceled. For airport/ # of GDPs canceled, the p -value is 0.0003 and for airport/ year/ # of GDPs canceled, the p -value is 0.0007. These results

suggest that there is a significant impact of CDM on the effectiveness of GDPs at various airports.

CDM went into full prototype operations in September 1998 at all airports. Analyses to measure GDP effectiveness were performed on “GDP Heavy” airports (airports having many planned GDPs) for three time intervals that are referred to as the “Pre-CDM”, “Transition to CDM” and “CDM” phases. The Pre-CDM Phase is from September 1996 through August 1997; The Transition to CDM Phase is from September 1997 through August 1998; The CDM Phase began in September 1998 and is currently being employed. The sources of the data for the analyses are the GDP Summaries and Logs from the ATCSCC and archived at Metron, Inc. (See Section 3.1 for a description of the data fields) The histograms in Figures 6.12 and 6.13 demonstrate the decrease in errors in the planning of GDPs (thus the increase in GDP effectiveness), but not the total elimination of planning errors.

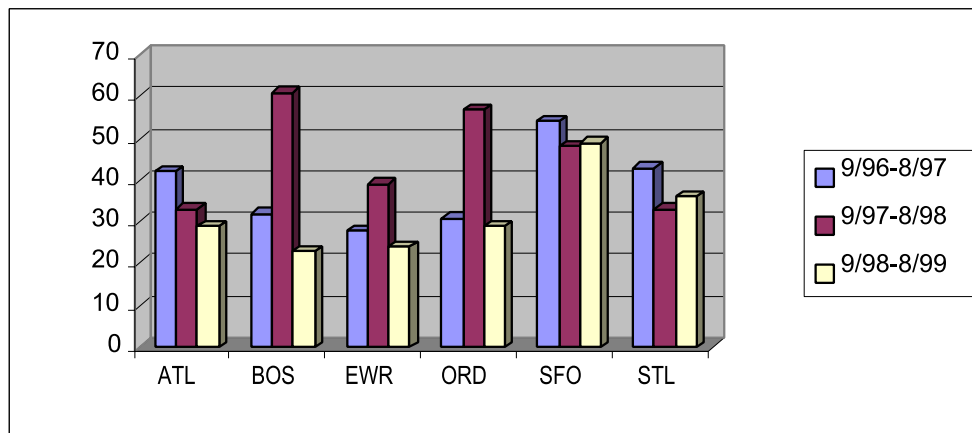


Figure 6.12: Percentage of GDPs Canceled

Figure 6.12 shows that, though, there is erratic behavior during the “Transition to CDM Phase”, the comparison of “pre-CDM” to “CDM” indicates that

there has been a decrease in the percentage of canceled GDPs. This metric is strongly affected by the weather conditions and forecasted demand. Cancellations of GDPs occur as soon as it is decided that there is no need for a GDP, which is sometimes based on better weather and demand information. Since there are instances when a cancellation of a GDP is a desired action, further stratification of cancellations is needed to extract those that are indeed undesirable. Therefore, a more representative metric of GDP effectiveness is the “Percentage of GDP Cancellations Near Start”. (See Figure 6.13)

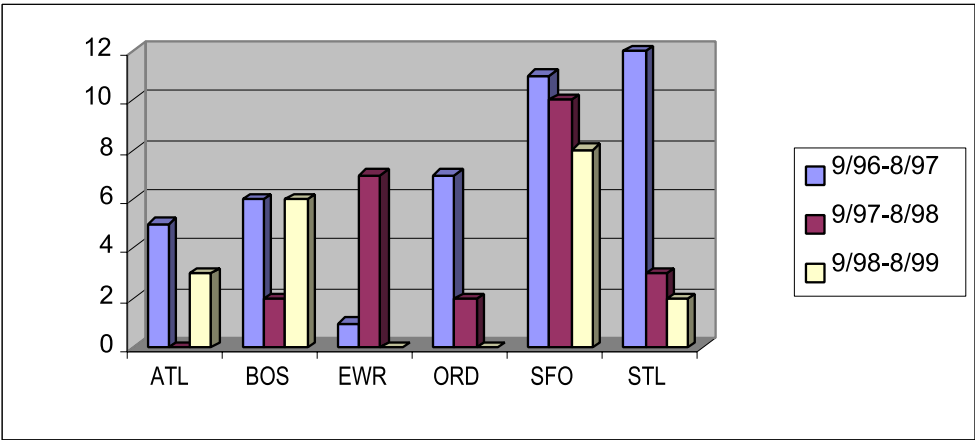


Figure 6.13: Percentage of GDP Cancellations Near Start

The type of GDP cancellation known as “GDP Cancellations Near Start” occurs within the time frame from 30 minutes prior to the planned start time of a GDP to 30 minutes after the planned start time of a GDP. In this context “start time of a GDP” refers to the time the first controlled flight is scheduled to land, thus referring to arrivals and not departures. Since ground delays are assigned and taken before flights take off, many flights will have absorbed delays well in advance of the start time of a GDP. If a GDP is canceled at or near its planned start time, this is an indication that there was no need for the GDP. Thus, nearly

all the (sometimes substantial) delays that have been absorbed prior to the start of the canceled GDP were found to be unnecessary. If there is a need for a cancellation of a GDP, then ideally it should be done in timely enough fashion to allow airlines to recover delay caused by the institution of the GDP. Under CDM, specialists have more accurate demand information to plan accordingly. Furthermore, they have the opportunity to delay the institution of a GDP until the “critical decision point”. Since they can delay it for a much longer period of time than during the “Pre-CDM” phase, they have the option to wait until better and more accurate weather and demand information is available. Thus, decisions to institute GDPs are based on superior data quality and accuracy and should generally result in fewer cancellations near GDP start times. It is interesting to note that out of all the “GDP Heavy” airports, San Francisco International Airport (SFO) seems to be most in need of better capacity estimations in order to further improve GDP effectiveness. According to Figure 6.12, the percentage of canceled GDPs for SFO still exceeds 40% even after the implementation of full prototype operations. What really compounds this observation is the fact that prototype operations were in effect at SFO since January 1998. It is evident in Figure 6.13 that CDM is making a significant positive impact on the percentage of GDPs canceled near the start time of the GDPs, yet SFO is the only airport with a percentage greater than 7%. It is due to these observations that the focus of this dissertation was on SFO and the development of models and capacity scenarios for SFO. The graphs in this chapter have shown that there is an increase in GDP effectiveness due to CDM. The models developed in this dissertation can be used in conjunction with CDM procedures to further improve GDP effectiveness.

6.5.2 Incremental GDPs

The results of Section 6.1 suggest that the Command Center should generally plan shorter programs (depending on the season or airborne delay cost factor). It was shown in Section 6.3.3 that planning shorter programs should lead to less total weighted delay. Though planning shorter programs would lead to a reduction in delay minutes, it may not lend itself to the operational preferences of the users (airlines). For example, if a GDP was initially planned to last 2 hours, the users would request either a temporary ground stop (GS) or increased miles-in-trail (MIT) restrictions.¹ Ground stops were described in Section 1.1. MIT is an initiative taken by the specialists that “spaces” out the aircraft more along routes in the airspace, in other words, a larger separation distance. The users would prefer a GS or MIT for shorter durations because they consider them to be less intrusive to their operations than a GDP. A short duration GDP would cause the loss of planning and substitution flexibilities. Using the current procedures, the biggest consequence of planning shorter programs, which may need to be revised/extended, is the loss of equity. In a short program, only flights that are close to the congested airport (short-haul flights) are given GD. Long haul flights are allowed to take off as originally scheduled. If the program needs to be revised/extended, the (long-haul) flights in the air will incur airborne delay, but the (short-haul) flights on the ground will incur extra GD to accommodate the airborne delay of the airborne flights that were not assigned any GD at all. Thus, short-haul flights are given an unequitable amount of delay. This is a serious issue that needs to be addressed.

¹Based on a phone conversation with Jill Charlton of the Quality Assurance Department at the ATCSCC on June 26, 2000.

One way to address the equity issue is to plan the short programs, but include both the short-haul and long-haul flights. Thus, if the GDP needs to be revised/extended, both types of flights would incur extra delay. This process is more equitable. Another way is to consider a GDP as incremental. In this case, at time t , decisions in a GDP are only made about flights scheduled to depart in the next time period ($t + 1$). This process is performed incrementally for each time period (only flights in the next time period are considered). This process is more equitable because all flights will be considered, just not at the same time period. In a given time period, it may be decided that no flights need to be delayed, but this depends on the weather conditions and not on the length of the flight (short-haul or long-haul).

Chapter 7

Conclusions

7.1 Contributions of Thesis

The goal of this thesis is to develop statistical models to estimate capacity probabilistic distribution functions (CPDFs) that contain capacity scenarios that are used as inputs into stochastic ground holding models. Towards the goal of this thesis, we estimated CPDFs by relative frequency histograms, which we generated using empirical GDP/weather data. The capacity scenarios called ACDs are derived from the CPDFs along with their respective probabilities. The structure of the 1-Parameter ACD, which is used to model morning fog at SFO, makes this a relatively simple process: the bins on the histogram of the CPDFs correspond to the duration of reduced capacity in the 1-Parameter ACD and the relative frequency of the bin gives the probability of the ACD.

Since weather is a continuous process, the CPDFs may vary in time. Set partitioning models are used to determine when a CPDF changes and, thus yields the breakpoints (boundaries) of the seasonal distributions. Various cost functions are presented and thoroughly explored. The outputs (sets of seasons) of the different cost functions are evaluated using a post-processing mean square ratio

criterion. The methodology used to determine and evaluate seasonal distributions has a broader applicability because it can be generalized to create a seasonal “clustering” technique that determines seasonal clusters while maintaining time-order of data.

The 1-Parameter ACDs with their associated probabilities were entered into the Hoffman-Rifkin (H-R) static stochastic ground holding model and resulted in a scenario that corresponded to one of the input scenarios. Analysis of the H-R model and its limitations resulted in our developing a new, simple, general decision model that would cost out all possible scenarios given a planned scenario. This general decision model allows for changes to be made to assigned ground delay according to dynamic changes in GDP plans. A procedure was provided for comparing proposed capacity scenarios to actual capacity scenarios. The modified H-R model developed in this thesis yielded less total weighted delay (sum of ground delay and weighted airborne delay) than the plans of the specialists at the ATCSCC for actual GDPs in 1998.

7.2 Directions for Further Research

With the tremendous growth of air traffic demand and congestion of the airspace comes an increase in the need for innovative methodologies and improved decision-support tools for effective CTFM. To this end, the models of this thesis make valuable contributions. The research in this thesis creates a foundation on which further research can be built for the development of models that would increase the effectiveness of CDM procedures.

Accurate estimation of capacity is contingent on the quality and accuracy of weather forecasts. The stochastic nature of weather can be captured by assuming

arrival capacity is a random variable described through arrival capacity distributions (ACDs). This thesis focused on modeling morning fog at SFO using the 1-Parameter ACD. To model other weather conditions at other major airports, the 2-Parameter ACDs can be used as described in Section 5.4. The techniques of this thesis can be directly applied to this case, although the size of the set partitioning problems will increase substantially.

Seasonal boundaries in Sections 5.3.3 and 5.3.4 are assumed to be fixed monthly boundaries. Since it is possible for seasons to begin or end on an arbitrary day, derivation of seasons with arbitrary boundaries may be worthwhile. These types of seasons may increase the value of the mean square ratio for contiguous seasons since the actual start and end days of weather seasons are captured.

A possible fruitful area of research is in sensitivity analysis of the input parameters of the H-R or modified H-R models. It may be worthwhile to determine the effects of altering the ground delay to airborne delay cost ratio or the probabilities of the input ACDs.

There should be a constant evaluation of the impacts of CDM on operational procedures and decision-making. For CDM performance analysis, we suggest that distributions of GDP durations prior to and after the inception of CDM be compared to ascertain the effects of CDM on the durations of GDPs. Distributions of GDP durations are affected by cancellations and revisions/extensions of GDPs. It is possible that the ability to revise a program either “up” or “down” results in longer-duration GDPs. Such an analysis would naturally employ the distribution of GDP durations obtained in this thesis.

Appendix A

Glossary

AAR: Airport Acceptance Rate

ACD: Arrival Capacity Distribution

AD: Airborne Delay

ADL: Aggregate Demand List

ANOVA: Analysis Of VAriance

AOC: Airline Operational Control Center

ARTA: Actual Runway Time of Arrival

ARTCC: Air Route Traffic Control Center

ARTD: Actual Runway Time of Departure

ASOS: Automated Surface Observing System

ATC: Airline Traffic Control

ATCSCC: Air Traffic Control System Command Center

ATL: Atlanta's Hartsfield Airport

BLUE: Best Linear Unbiased Estimators

BOS: Boston's Logan Airport

CDF: Cumulative Distribution Function

CDM: Collaborative Decision Making

CMA: Centered Moving Average
CPDF: Capacity Probabilistic Distribution Function
CR: Collaborative Routing
CTA: Controlled Time of Arrival
CTD: Controlled Time of Departure
CTFM: Collaborative Traffic Flow Management
df: degrees of freedom
EDF: Empirical Distribution Function
ETA: Estimated Time of Arrival
ETD: Estimated Time of Departure
EWR: Newark's Airport
FAA: Federal Aviation Administration
FADE: FAA/Airline Data Exchange
FSM: Flight Schedule Monitor
GD: Ground Delay
GDP: Ground Delay Program
GDP-E: Ground Delay Program Enhancements
GHP: Ground Holding Problem
GL: GDP Length
GS: Ground Stop
H-R: Hoffman-Rifkin (static stochastic ground holding model)
IFR: Instrument Flight Rules
IP: Integer Program
KS: Kolmogorov-Smirnov
LP: Linear Program

MA: Moving Average
MIT: Massachusetts's Institute of Technology
MIT: Miles-In-Trail
MS: Mean Square
NAS: National Airspace System
NCDC: National Climatic Data Center
NEXTOR: National center of EXcellence for AviaTion Operations Research
NP: Nondeterministic Polynomial
NWS: National Weather Service
OAG: Official Airline Guide
OETA: Original Estimated Time of Arrival
OETD: Original Estimated Time of Departure
OLS: Ordinary Least Squares
ORD: Chicago's O'hare Airport
PAAR: Planned Airport Acceptance Rate
PDF: Probability Distribution Function
RBS: Ration By Schedule
SFO: San Francisco's International Airport
SS: Sum of Squares
STL: St. Louis' International Airport
TAD: Total Airborne Delay
TFM: (air) Traffic Flow Management
TFMP: Traffic Flow Management Problem
TGD: Total Ground Delay
TU: Totally Unimodular

TWD: Total Weighted Delay

VAPS: Visual APproacheS

VFR: Visual Flight Rules

WLS: Weighted Least Squares

Appendix B

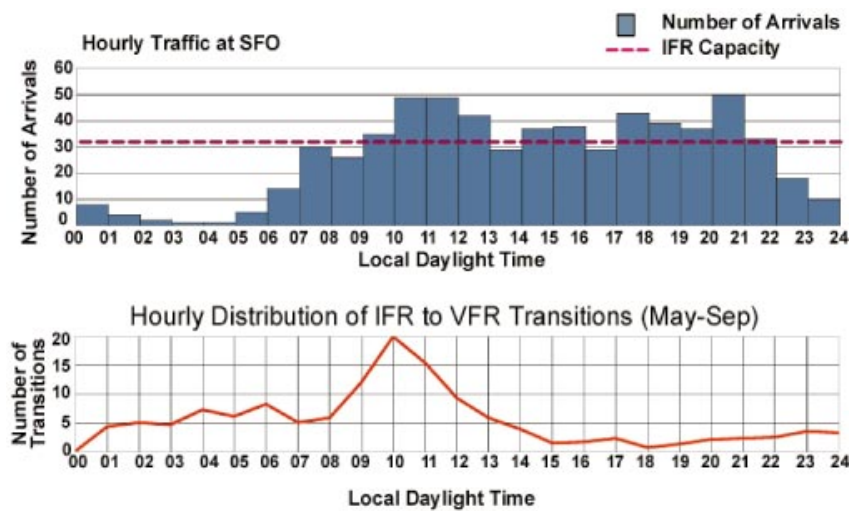
MIT's Lincoln Laboratory Graphs and Chapter 5 Tables

The following graphics are being reproduced with permission of David A. Clark of MIT's Lincoln Laboratory. Graphics are adapted from :

Clark, D.A. (1995), "Characterizing the Causes of Low Ceiling and Visibility at U.S. Airports", *6th Conference on Aviation Weather Systems*, Dallas, TX.



SFO Operations and Cloud Climatology



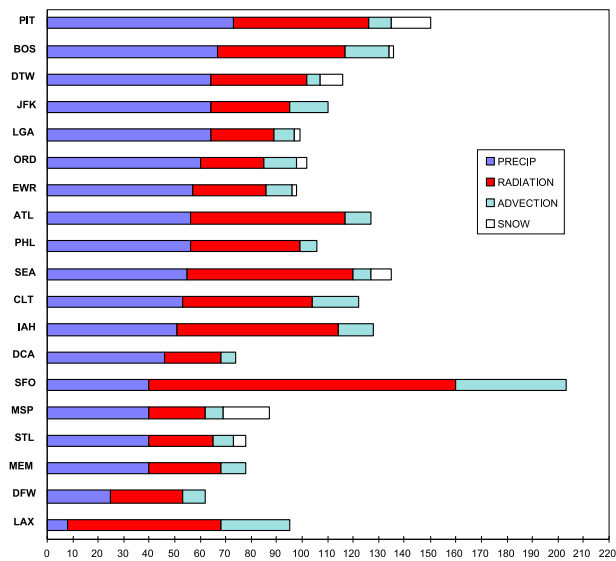
Weather_Web-1
Group 43 3/6/00

MIT Lincoln Laboratory

Figure B.1: IFR Capacity at SFO



Annual IFR Event Frequency by Cause



MIT Lincoln Laboratory

Weather_Web-3
Group 43 3/6/00

Figure B.2: Frequency of IFR events at Major US Airports

133 Seasons:	SoSqs		No Singles	Normalized		Variances
N=3 (Cov)	Apr-Nov		Apr-Nov	Apr-Nov		Jan-Dec
	Dec/Jan		Dec/Jan	Dec/Jan		Jul
	Feb/Mar		Feb/Mar	Feb/Mar		Jul/Aug
N=3 (Par)	Apr-Nov		Apr-Nov	Apr-Nov		Jul
	Dec/Jan		Dec/Jan	Dec/Jan		Aug
	Feb/Mar		Feb/Mar	Feb/Mar		Sep-Jun
N=4 (Cov)	Apr-Nov		Apr-Nov	Apr-Oct		Jan-Dec
	Dec		Dec/Jan	Nov		Jul
	Jan		Feb/Mar	Dec/Jan		Jul/Aug
	Feb/Mar		Jul/Aug	Feb/Mar		Jul/Sep
N=4 (Par)	Apr-Nov		Apr-Sep	Apr-Oct		Jul
	Dec		Oct/Nov	Nov		Aug
	Jan		Dec/Jan	Dec/Jan		Sep
	Feb/Mar		Feb/Mar	Feb/Mar		Oct-Jun
N=5 (Cov)	Apr-Nov		Apr-Nov	Apr-Oct		Jul
	Dec		Apr/May	Nov		Aug
	Jan		Dec/Jan	Dec		Sep
	Feb		Feb/Mar	Jan		Oct
	Mar		Jul/Aug	Feb/Mar		Nov-Jun
N=5 (Par)	Apr-Nov		Apr-Jun	Apr-Oct		Jul
	Dec		Jul-Sep	Nov		Aug
	Jan		Oct/Nov	Dec		Sep
	Feb		Dec/Jan	Jan		Oct
	Mar		Feb/Mar	Feb/Mar		Nov-Jun

Table B.1: Set Covering and Partitioning Solutions of GDP Seasons(n=133)

For 133 Possible Seasons:	N = 3	N = 4	N = 5
Covering (GDP Data)	<div>Apr-Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div>	<div>Apr-Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div> <div>Jul/Aug</div>	<div>Apr-Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div> <div>Jul/Aug</div> <div>Oct/Nov</div>
Partitioning (GDP Data)	<div>Apr-Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div>	<div>Apr-Sep</div> <div>Oct/Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div>	<div>Apr-Jun</div> <div>Jul/Aug</div> <div>Sep-Nov</div> <div>Dec/Jan</div> <div>Feb/Mar</div>
Covering (Weather Data)	<div>Apr-Nov</div> <div>Dec-Mar</div> <div>May/Jun</div>	<div>Apr-Nov</div> <div>Nov-Feb</div> <div>Mar/Apr</div> <div>May/Jun</div>	<div>Apr-Nov</div> <div>Dec/Jan</div> <div>Jan/Feb</div> <div>Mar/Apr</div> <div>May/Jun</div>
Partitioning (Weather Data)	<div>May/Jun</div> <div>Jul-Sep</div> <div>Oct-Apr</div>	<div>Mar/Apr</div> <div>May/Jun</div> <div>Jul-Sep</div> <div>Oct-Feb</div>	<div>Mar/Apr</div> <div>May/Jun</div> <div>Jul-Sep</div> <div>Oct/Nov</div> <div>Dec-Feb</div>

Table B.2: Set Covering/Partitioning Solutions using KS statistics (n=133)

Appendix C

Chapter 5 Graphs

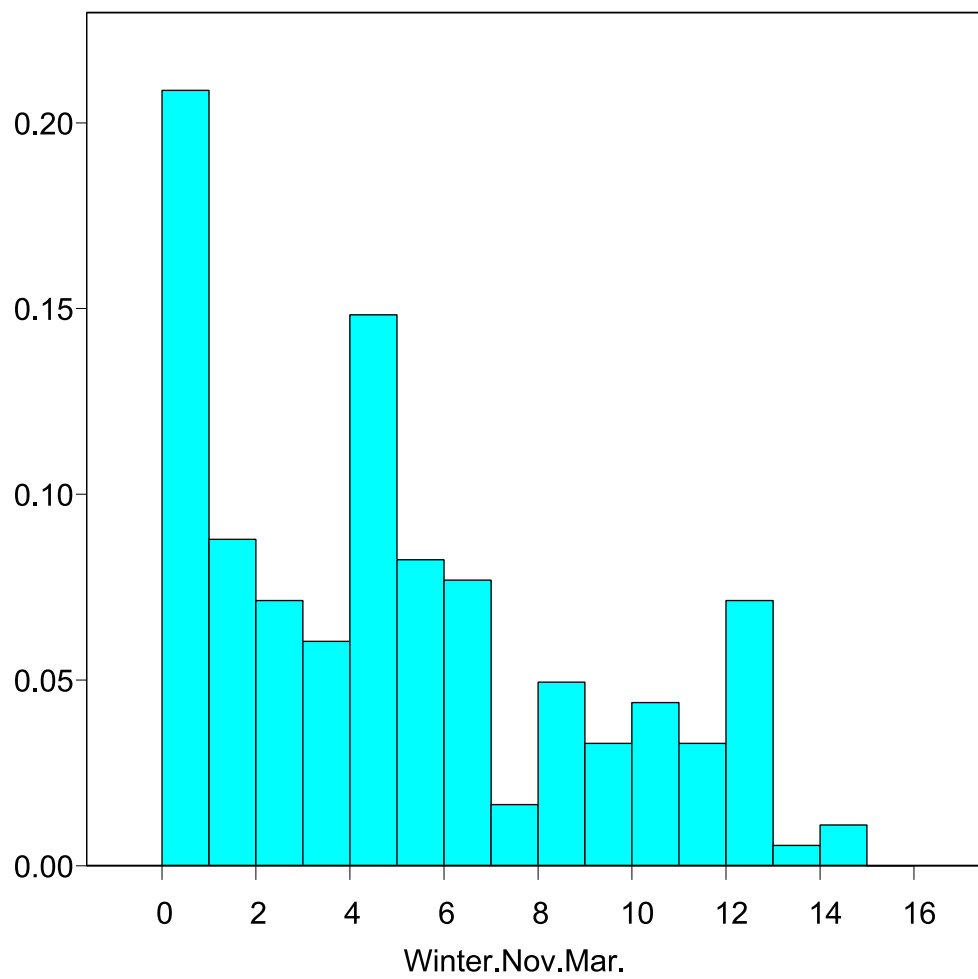


Figure C.1: Relative Frequency Histogram for Winter GDP Season

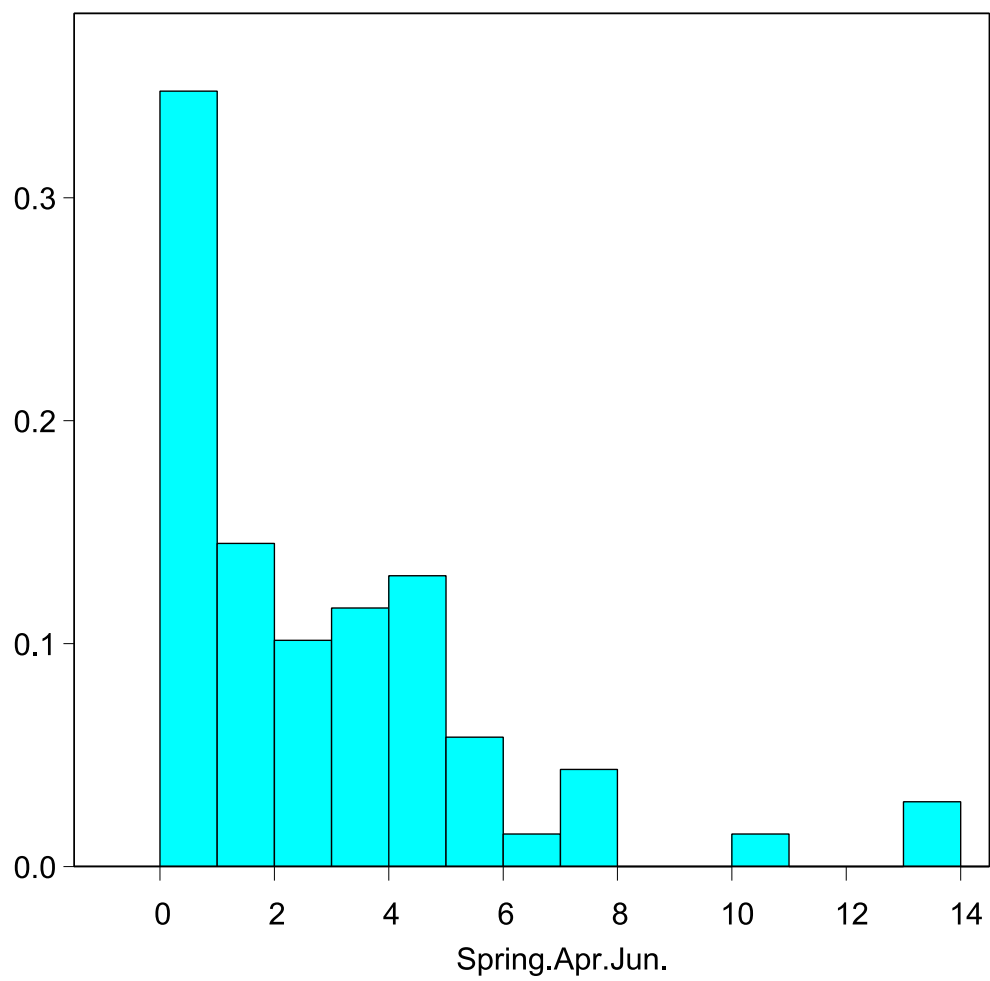


Figure C.2: Relative Frequency Histogram for Spring GDP Season

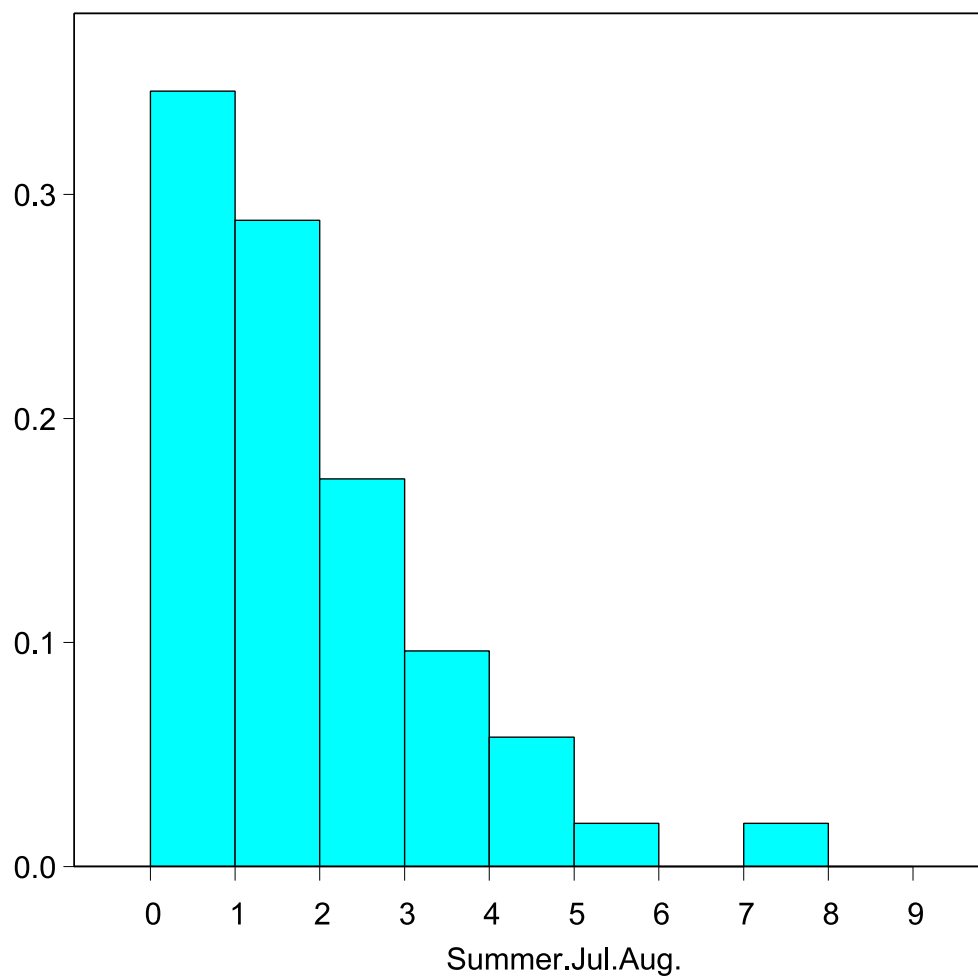


Figure C.3: Relative Frequency Histogram for Summer GDP Season

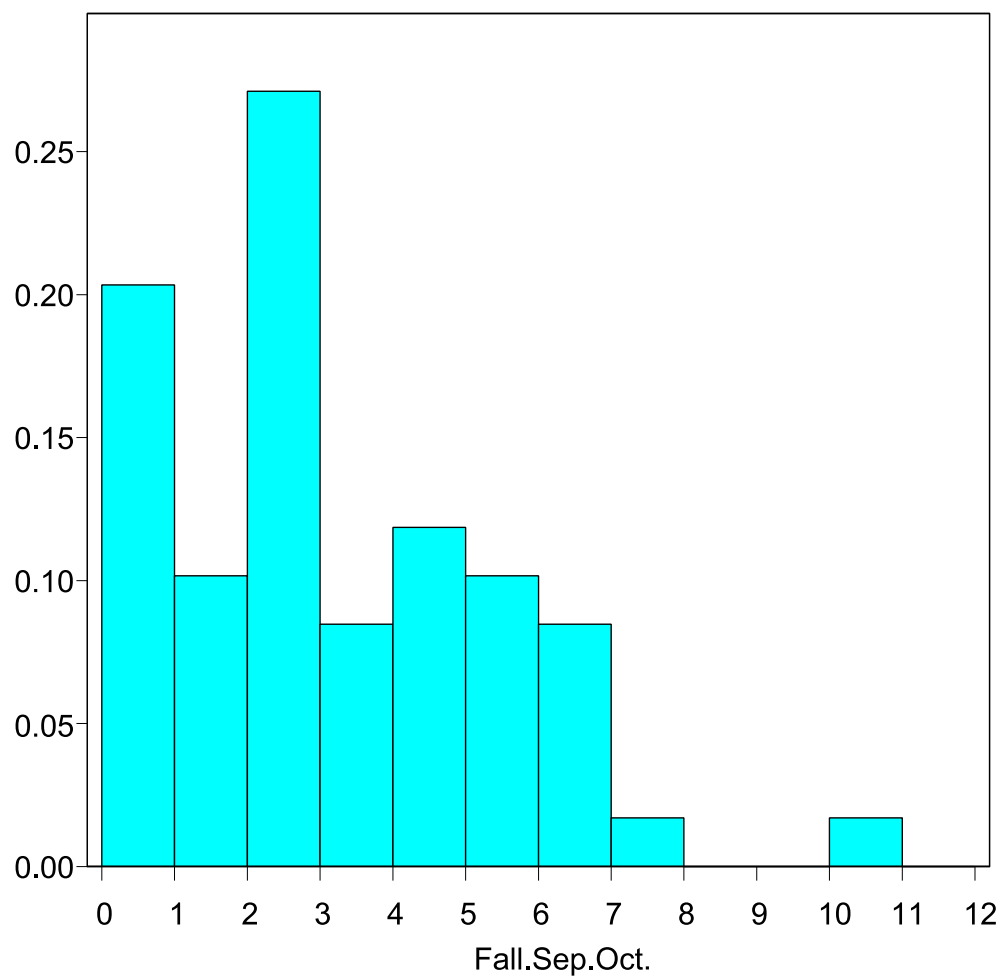


Figure C.4: Relative Frequency Histogram for Fall GDP Season

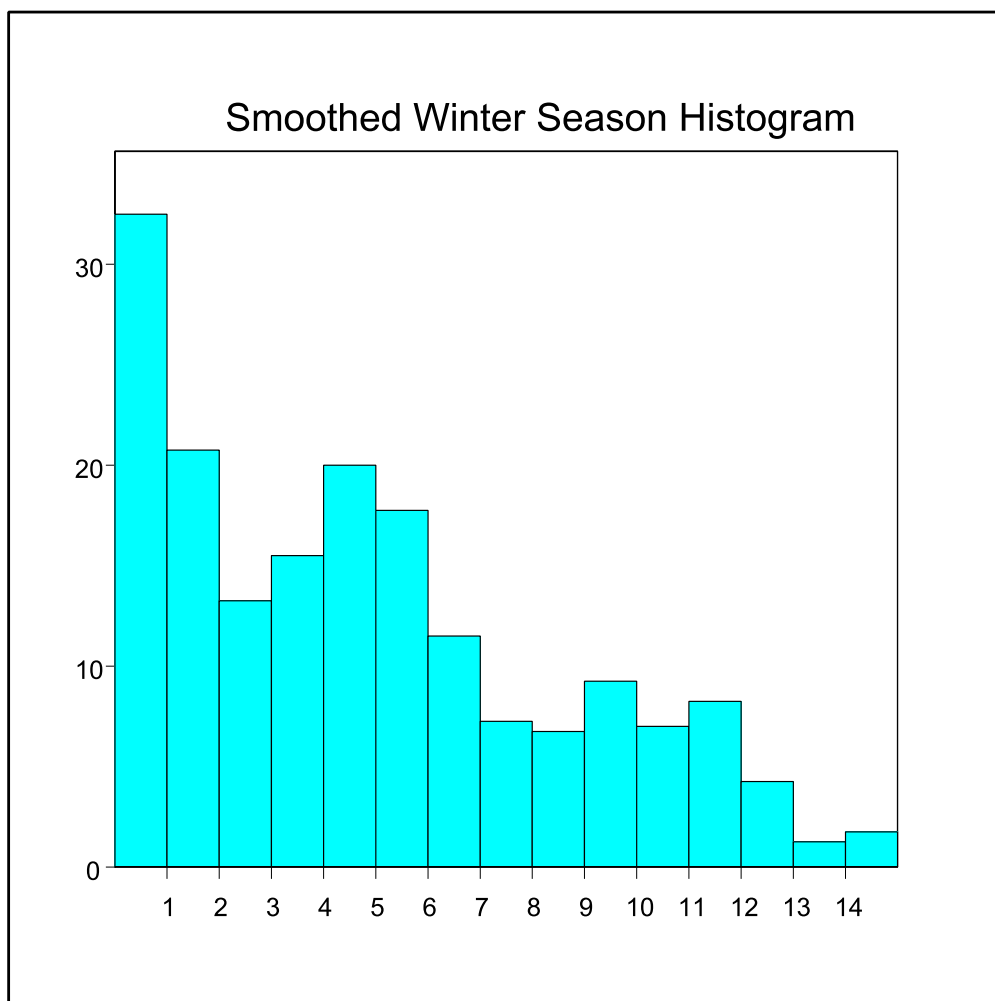


Figure C.5: Smoothed Histogram for Winter GDP Season

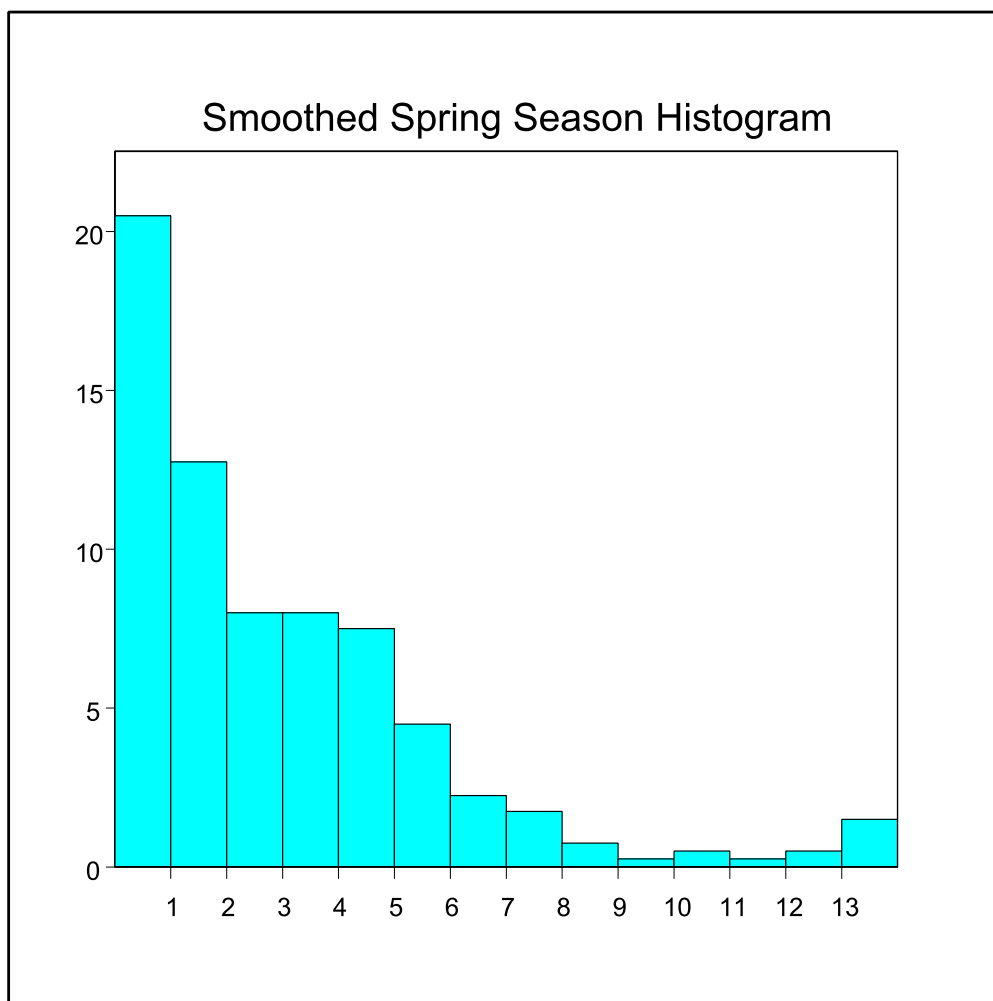


Figure C.6: Smoothed Histogram for Spring GDP Season

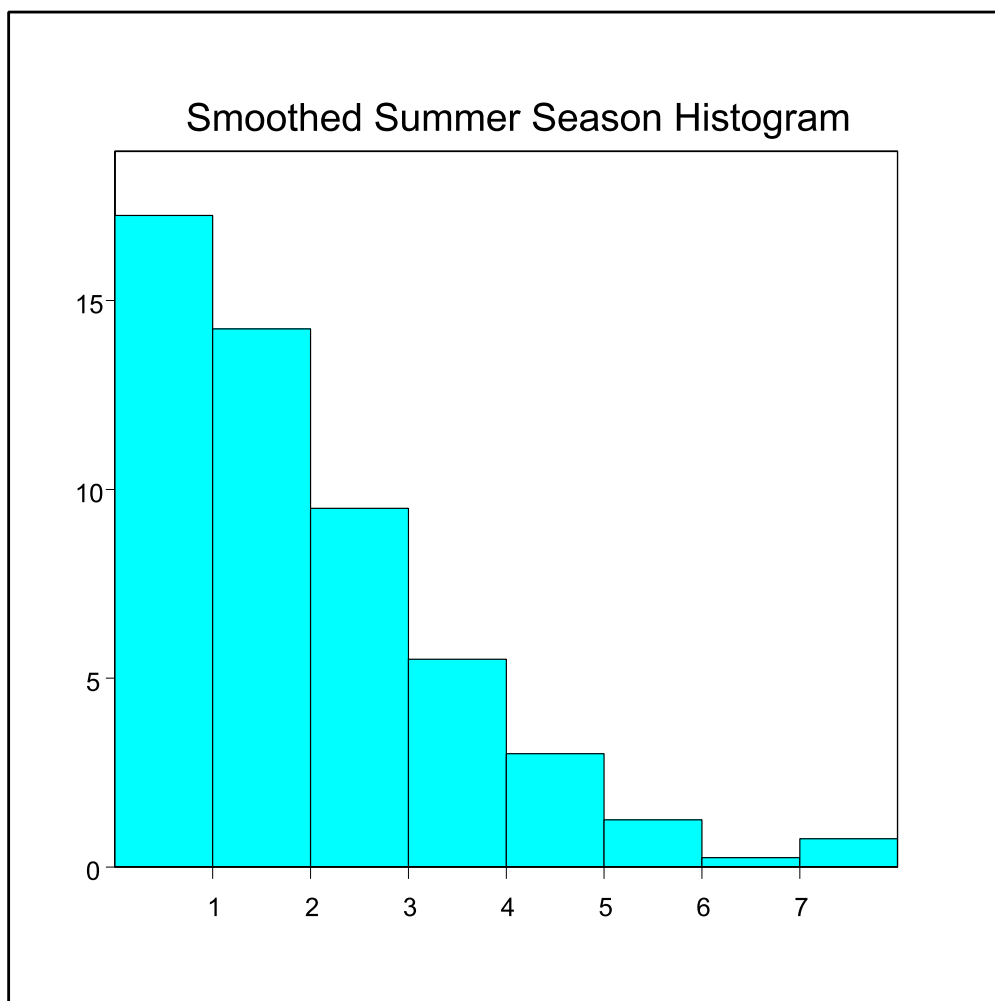


Figure C.7: Smoothed Histogram for Summer GDP Season

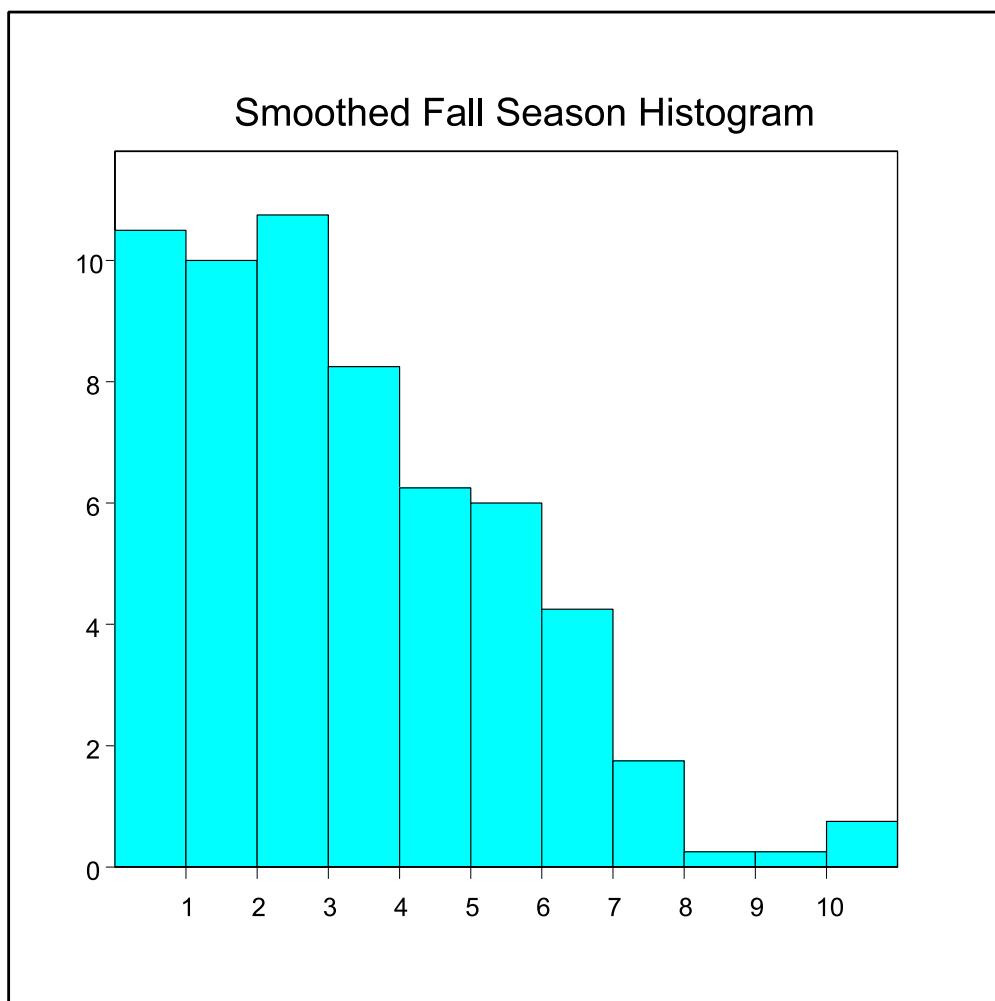


Figure C.8: Smoothed Histogram for Fall GDP Season

Appendix D

Comparison of Results of H-R Model (M_PAAR) and Command Center Plan (CC_PAAR)

BIBLIOGRAPHY

- [1] “Aircraft Operating Costs by Stage of Flight and Associated Air Traffic Control Delay Costs Based on a Typical Flight,” 1999, *Air Transport Association*, Washington, D.C., 2000.
- [2] Andreatta, G., and Romanin-Jacur, G. (1987) “Aircraft Flow Management Under Congestion,” *Transportation Science*, **21**, 249-253.
- [3] Andreatta, G., Odoni, A.R. and Richetta, O. (1993) “Models for the Ground-Holding Problem,” in *Large-Scale Computation and Information Processing in Air Traffic Control*, L. Bianco and A.R. Odoni, eds., Springer-Verlag, Berlin, 125-168.
- [4] Ball, M., Hoffman, R., Hall, W. and Muharremoglu, A. (1998) “Collaborative Decision Making in Air Traffic Management: A Preliminary Assessment,” Technical Report RR-99-3, NEXTOR, UC Berkeley.
- [5] Ball, M., Hoffman, R., Odoni, A., and Rifkin, R. (1999), “The Static Stochastic Ground Holding Problem with Aggregate Demands,” Technical Report RR-99-1, NEXTOR, UC Berkeley.
- [6] Ball, M., Chen, C-Y., Hoffman, R., Vossen, T. (1999) “Collaborative Decision Making in Air Traffic Management: Current and Future Research

- Directions,” *Proceedings of ATM '99, Workshop on Advanced Technologies and Their Impact on Air Traffic Management in the 21st Century*.
- [7] Bertsimas, D. and Stock, S. (1998) “The Air Traffic Flow Management Problem with Enroute Capacities”, *Operations Research*, **46**, 406-422.
 - [8] Birge, J.R. and Louveaux, F. (1997) *Introduction to Stochastic Programming*, Springer-Verlag, New York.
 - [9] Bodin, L.D. (1972) “A Graph Theoretic Approach to the Grouping of Ordering Data,” *Networks*, **2**, 307-310.
 - [10] Chatfield, C. (1984) *The Analysis of Time Series: An Introduction*, 3rd Edition, Chapman and Hall, New York.
 - [11] Clark, D.A. and Wilson, F.W. (1996) “The Marine Stratus Initiative at San Francisco International Airport”, *Project Report ATC-252*, MIT Lincoln Laboratory, Lexington, MA.
 - [12] Cody, R.P. and Smith, J.K. (1997) *Applied Statistics and the SAS Programming Language* (4th ed.), Prentice Hall, New Jersey.
 - [13] Draper, N.R. and Smith, H. (1981) *Applied Regression Analysis*, John Wiley and Sons, Inc., New York.
 - [14] Everitt, B. (1994) *A Handbook of Statistical Analysis Using S-Plus*, Chapman and Hall, New York.
 - [15] Fulkerson, D.R. and Gross, D.A. (1965), “Incidence Matrices and Interval Graphs,” *Pacific Journal of Mathematics*, **15**, 833-835.

- [16] Gilbo, E.P. (1993) "Airport Capacity: Representation, Estimation, Optimization," *IEEE Transactions on Control Systems Technology*, **1**, 308-313.
- [17] Gilbo, E.P. (1997) "Optimizing Airport Capacity Utilization in Air Traffic Flow Management Subject to Constraints at Arrival and Departure Fixes," *IEEE Transactions on Control Systems Technology*, **5**, 490-503.
- [18] Glockner, G.D. (1996) "Effects of Air Traffic Congestion Delays Under Several Flow Management Policies," Ph.D. Dissertation, Georgia Institute of Technology.
- [19] Hall, W. (1999) "Efficient Capacity Allocation in a Collaborative Air Transportation System," Ph.D. Dissertation, Massachusetts Institute of Technology.
- [20] Hartigan, J.A. (1975) *Clustering Algorithms*, John Wiley and Sons, New York.
- [21] Hoffman, R. (1997) "Integer Programming Models for Ground-Holding in Air Traffic Flow Management," Ph.D. Dissertation, University of Maryland.
- [22] Hoffman, R., Ball, M., Hall, B., Odoni, A.R. and Wambsganss, M. (1999) "Collaborative Decision Making in Air Traffic Flow Management," Technical Report RR-99-2, NEXTOR, UC Berkeley.
- [23] Hogg, R.V. and Craig, A.T. (1995) *Introduction to Mathematical Statistics*, 5th Edition, Prentice Hall, Englewood Cliffs, New Jersey.
- [24] Knorr, D., Wetherly, J., Ball, M., Hoffman, R., Inniss, T., Tanino, M., Shisler, L., Brennan, M., Blucher, M., DeArmon, J., Darrow, A.M., Howard,

- K., Gilbo, E., Danz, R., Biros, J. (1999) “An Operational Assessment of Collaborative Decision Making in Air Traffic Management: Measuring User Impacts through Performance Metrics,” Document Control Number: R90145-01, FFP1, Federal Aviation Administration.
- [25] Kollman, K., Kupper, K. and Wetherly, J. (1997) “Collaborative Decision Making in Aviation Transportation”, <http://www.metsci.com/cdm>.
- [26] MathSoft, Inc. (1997) *S-PLUS Guide to Statistics*, Version 4, Seattle, Washington.
- [27] MacDonald, L. (1998) “Collaborative Decision Making in Aviation,” *Journal of Air Traffic Control*, April-June 1998.
- [28] Nemhauser, G.L. and Wolsey, L.A. (1988) *Integer and Combinatorial Optimization*, John Wiley and Sons, Inc., New York.
- [29] Neter, J. and Wasserman, W. (1974) *Applied Linear Statistical Models*, Richard D. Irwin, Inc., Homewood, Illinois.
- [30] Newell, G.F. (1979) “Airport Capacity and Delays,” *Transportation Science*, **13**, 201-241.
- [31] Odoni, A.R. (1987) “The Flow Management Problem in Air Traffic Control,” In *Flow Control of Congested Networks*, A.R. Odoni, L.Bianco, and G. Szego, eds., Springer-Verlag, Berlin, 269-288.
- [32] Richetta, O. (1991) “Ground Holding Strategies for Air Traffic Control under Uncertainty,” Ph.D. Dissertation, Massachusetts Institute of Technology.

- [33] Richetta, O. and Odoni, A.R. (1993) "Solving Optimally the Static Ground-Holding Policy Problem in Air Traffic Control," *Transportation Science*, **27**, 228-238.
- [34] Richetta, O. (1995) "Optimal Algorithms and a Remarkably Efficient Heuristic for the Ground-Holding Problem in Air Traffic Control," *Operations Research*, **43**, 758-770.
- [35] Rifkin, R. (1998) "The Static Stochastic Ground-Holding Problem," Master's Thesis, Massachusetts Institute of Technology.
- [36] Rohatgi, V.K. (1984) *Statistical Inference*, John Wiley and Sons, New York.
- [37] Saigal, R. (1968) "A Constrained Shortest Route Problem," *Operations Research*, **16**, 205-209.
- [38] The SAS Institute Inc. (1990) *SAS/STAT User's Guide*, Version 6, 4th Edition, Cary, NC.
- [39] Scheffe', H. (1959) *The Analysis of Variance*, John Wiley and Sons, New York.
- [40] Terrab, M. (1990) "Ground Holding Strategies for Air Traffic Control", Ph.D. Dissertation, Massachusetts Institute of Technology.
- [41] Terrab, M. and Odoni, A.R. (1993) "Strategic Flow Management for Air Traffic Control," *Operations Research*, **41**, 138-152.
- [42] Vanchieri, A. (1997) "Aviation Dream Team," *ORMS Today*, April, 39-41.
- [43] Venables, W.N. and Ripley, B.D.(1994), *Modern Applied Statistics with S-Plus*, Springer-Verlag, New York.

- [44] Venkatakrisnan, C., Barnett, A. and Odoni, A.R. (1993) "Landings at Logan Airport: Describing and Increasing Airport Capacity," *Transportation Science*, **27**, 211 - 227.
- [45] Vranas, P., Bertsimas, D. and Odoni, A.R. (1994) "The Multi-Airport Ground-Holding Problem in Air Traffic Control," *Operations Research*, **42**, 249-261.
- [46] Vranas, P., Bertsimas, D. and Odoni, A.R. (1994) "Dynamic Ground-Holding Policies for a Network of Airports," *Transportation Science*, **28**, 275-291.
- [47] Wambsganss, M. (1996) "Collaborative Decision Making Through Dynamic Information Transfer," *Air Traffic Control Quarterly*, **4**, 107-123.
- [48] Wang, H. (1991) "A Dynamic Programming Framework for the Global Flow Control Problem in Air Traffic Management," *Transportation Science*, **25**, 308-313.
- [49] Wilson, F.W., and Clark, D.A. (1997) "Estimation of the Cost of Ceiling and Visibility Events at Major Airports," *7th Conference on Aviation , Range, and Aerospace Meteorology*, Long Beach, CA.
- [50] Wolsey, L.A. (1998) *Integer Programming*, John Wiley and Sons, Inc., New York.